

Whistleblowers in peril

The US Congress should reverse a pernicious removal of protection of federal employees.

In another worrying instance of its tendency to quietly arrogate new powers to itself, the Bush administration has reversed two decades of precedent and declared that important whistleblower protections in the Clean Water Act do not apply to federal workers.

This binding change in the interpretation of the law was instigated by a top Department of Justice lawyer a year ago. Steven Bradbury, acting assistant attorney-general in the department's Office of Legal Counsel, gave a straightforward reason for his decision: the Clean Water Act does not list the US government as a 'person' in its definition of employers from whom whistleblowers may seek redress in the event of retaliation by their bosses. He concluded that the ancient legal doctrine of sovereign immunity — which says the government must explicitly consent to be sued — makes the federal government immune from the whistleblower provisions in the law.

The water law was written to protect workers in the private and public sectors who report breakdowns in its enforcement, manipulations of science, or clean-up failures. Its whistleblower provisions essentially apply to any action a worker might take in a sincere effort to do a good job — and hence go further than a different, government-wide whistleblower law that is still in place but that protects only the reporting of gross mismanagement or violations of law.

The water law's provisions have real teeth: whistleblowers who are found to have legitimate complaints are eligible for reinstatement to lost jobs, back pay, and compensatory damages for loss of reputation and emotional distress — damages that in the past have ranged in the tens of thousands of dollars.

Exempting federal employees would expose to retaliation some 170,000 members of the federal workforce in a dozen different agencies, from the Forest Service to the US Geological Survey, who might make efforts in good faith to see that the law is properly enforced. Scientists could feel this particularly strongly, as the problems they encounter — such as the skewing of a methodology or the removal of a conclusion from a report — don't typically violate a law. This

change is bound to suppress their willingness to report such events.

Superficially, the justice department has made a defensible case. But it goes against two decades of precedent during which the Department of Labor adjudicators charged with administering whistleblower law repeatedly rejected arguments for the government's sovereign immunity under the water law. Legal doctrine holds that, when an agency such as the labour department has a long-standing interpretation of a law, as in this instance, and Congress does nothing to change it, it can be assumed that Congress accepts that interpretation.

But what is particularly disturbing about this change is the way it was brought in under the radar, remaining unpublished for 12 months and unknown to the federal workers potentially affected by it. It only became public last week, when the advocacy group Public Employees for Environmental Responsibility released a letter from Bradbury, which it obtained under the Freedom of Information Act after stumbling upon a reference to it in a whistleblower complaint. This is hardly a fitting approach to jurisprudence in a purportedly open and democratic society.

It is common knowledge that the Bush administration has fought against implementing more stringently protective environmental laws, and its enforcement of existing laws has been weak to a fault; according to the justice department's figures, government requests for criminal prosecutions of environmental lawbreakers fell by half in the five years to 2005. In such an atmosphere, whistleblowers within the government become a key defence against further erosion of environmental standards. Removing their protections seems all but certain to hasten this erosion.

There is a possible remedy, however: Congress should amend the Clean Water Act to define the US government as an employer against whom whistleblower complaints can be brought. ■

"What is particularly disturbing about this change is the way it was brought in under the radar."

Flickers of consciousness

Brain scans are forcing researchers to reconsider the state of outwardly unresponsive patients.

The news that a patient in a vegetative state has shown signs of awareness of the outside world, and an ability to perform mental tasks on request (see page 132), marks a turning point in the investigation of these enigmatic medical conditions. It also means that neuroscientists have some serious thinking to do.

The evaluation of vegetative states, in which patients are often 'awake' but show no outward sign that they are aware of themselves or their environment, has always been difficult, simply because of the

huge variability in the nature and severity of the brain injuries that underpin them. What is often so heartbreaking for friends and relatives is that the patient's mental quality of life is unknowable.

In the wake of the latest research (A. M. Owens *et al.* *Science* **313**, 1402; 2006), the inevitable calls for all 'vegetative' patients to be considered internally conscious should not be heeded. This case has little bearing on ethical questions over whether or not the most severe cases should be denied treatment and allowed to die. Clearly, for those with the bleakest prognosis — those who have shown no improvement over a period of at least 12 months and are categorized as being in a 'permanent vegetative state' — the chances of finding any sign of awareness are negligible.

But when considering those at the less severe end of the spectrum, such as the British patient at the heart of the latest research, the new



discovery presents a conundrum. Loved ones of similar patients will clamour for a chance to see whether they, too, are capable of similar feats. Clinical neuroscientists should stress that no two vegetative patients are alike, and that this remarkable case may be a one-off.

Unique or not, the discovery blurs the distinction between a vegetative patient and those classed as 'minimally conscious' — showing limited or intermittent interaction with their surroundings. As the patient has reportedly improved since the study was carried out, it is entirely possible that the brain scans simply reflected early signs of her recovery. But with brain scanning of vegetative patients becoming more common (although far from routine), the medical criteria for diagnosing a vegetative condition will have to be scrutinized.

A patient is currently classed as vegetative if there are no outward signs of genuine awareness, rather than simple reflex responses to

stimuli such as pain. But although this patient remained inscrutable throughout the study, her brain scans suggested she was picturing herself scampering across a tennis court, or roaming around her home. If such rich internal mental processes are seen in other outwardly vegetative patients, then clinicians should revise their opinion on whether or not these patients are really vegetative at all.

Some clinicians take issue with the use of the word 'vegetative' and the unpleasant imagery it evokes. Replacing it with the term 'outwardly unresponsive' would help to eliminate any confusion over whether internal awareness should be factored into a diagnosis. But the ultimate impact of this discovery will hinge on whether more patients are found to have this sort of inner responsiveness. Then, the issue of what constitutes consciousness in these patients — a question that transcends mere brain imaging — will come alive. ■

Resisting terrorism

Government agencies, stakeholders and fellow researchers need to support primate researchers.

A neuroscientist at the University of California, Los Angeles (UCLA), last month declared that he would abandon his federally funded research on non-human primates after being terrorized for years by opponents of animal research. This is only the latest of many such incidents in the United States and Europe. Yet it should serve as a wake-up call to the US research community as a whole and those who have a stake in it.

The tactics of extremist groups go beyond expressing their opinions via peaceful protest. They include intimidating scientists' families, and scaring others to the point where they are reluctant to speak out for fear of being targeted themselves. Some are not above using terror tactics — in the recent UCLA case, a bomb was left on the doorstep of a researcher's 70-year-old neighbour. The strategy is often to single out vulnerable researchers for harassment. Neuroscientists who do basic research on primates are deemed attractive targets because of the relative ease of whipping up public sympathy for the animals and of portraying the research as having no practical benefit.

However, the problem is not specific to primate studies or basic research. There is clear evidence that extremists, emboldened by their success, are pursuing the end of all animal research. Researchers should not complacently imagine that this is someone else's problem just because their own lab does no primate research.

It is essential that scientists, scientific societies, university administrators, government and law-enforcement agencies coordinate their responses to this threat to fundamental research. Universities and scientific agencies have a responsibility to stand behind their researchers in their efforts to conduct studies that have already been stringently vetted through review both by the scientific community for its scientific value, and by animal-care committees. In the United States, the latter include veterinary surgeons as well as local-community representatives with no vested interest in the research. They are charged with ensuring ethical standards and minimizing animal discomfort.

Individual scientists are vulnerable to the enormous pressure that

well-organized groups can apply and are limited in how they can respond. Universities and scientific agencies can and should exert their muscle to provide support to targeted scientists, whether it be through material support, defending the value of their research to the public on their behalf, or by working with government and law-enforcement agencies to protect them.

Scientists in Britain — no strangers to animal activism — have benefited from recent legislation amending the Serious Organized Crime Act, which defines animal extremism as terrorism. In the United States, people should familiarize themselves with (and consider supporting) similar bills, such as the Animal Enterprise Terrorism Act currently going through Congress, which would help to protect researchers from acts of intimidation and violence.

At the same time, neuroscientists who work with primates should recognize that although the value of primate research may seem so obvious as not to require explanation, it is often not recognized as such, even by other biologists.

Non-human primates are still uniquely valuable in animal research because of the similarity of their brains and many of their behaviours to ours. Non-invasive research methods in humans such as functional magnetic research imaging (fMRI) — often held up by opponents of primate research as a technology that obviates invasive procedures — can provide a complementary and more indirect view, but they do not replace the information that can be gained through more direct manipulations.

These provide the means to crack some basic questions about the neural computations, circuitry and organization that underlie human behaviour. Such information can also directly or indirectly aid research that could have practical applications. For example, the development of deep brain stimulation as a treatment of Parkinson's disease has benefited enormously not just from a primate model for the disease, but also from an understanding of the basic principles of basal-ganglia circuitry worked out through anatomy, lesion and recording studies.

The value and promise of such knowledge is clear. The researchers who pursue it need to do more to explain to the public and their colleagues why primate experiments are necessary, and to be sensitive to concerns about animal welfare. They require and deserve the coherent support of their fellow researchers and stakeholders. ■

RESEARCH HIGHLIGHTS

Monsoon forecasting

Science doi:10.1126/science.1131152 (2006)

Detailed measurements of sea surface temperatures may improve predictions of India's monsoon, say Martin Hoerling of the National Oceanic and Atmospheric Administration in Boulder, Colorado, and his colleagues.

The crop-watering rains are known, from historical records, to be prone to fail when a climatic event known as El Niño warms the surface waters of the Pacific. Hoerling and his team compared El Niño events that led to drought with those that did not. They conclude that the risk of monsoon failure depends on where the waters warm most: elevated temperatures in the central Pacific seem to be more strongly associated with drought than elevated temperatures in the eastern Pacific. Better predictions of monsoon failure would allow rural communities to prepare in advance.



STRDEL/AP/GETTY

NANOTECHNOLOGY

Short straws

Nano Lett. doi:10.1021/nl061786n (2006)

An approach to building complex pillars from polymers and gold, devised by chemists at Harvard University in Cambridge, Massachusetts, could find application in solar cells and light-emitting diodes.

George Whitesides and his team used the pores of an aluminium oxide membrane as a template to make arrays of columns some 200 nanometres wide and a few micrometres long. They produced an assortment of composite structures: some with a hollow polymer core surrounded by a gold shell electrode, others containing adjoining gold and polymer segments. The researchers suggest it would be possible to create efficient light-collecting or light-emitting devices from these structures by filling their hollow parts with a conducting polymer.

EVOLUTIONARY GENETICS

One small hop for a gene...

Science 313, 1448–1450 (2006)

Genes relocating from one chromosome to another could underlie a defining event in the origin of species, say researchers.

A new species arises when populations within an existing species have diverged so far that their members can no longer interbreed to produce fertile offspring. John P. Masly of the University of Rochester, New York State, and his colleagues identify a novel molecular mechanism by which this can happen.

They identified a gene essential for male fertility, *JYAlpha*, on the third chromosome of the fruitfly *Drosophila simulans*, but

on the fourth chromosome of *Drosophila melanogaster*. *JYAlpha* hopped between the chromosomes at least 300,000 years ago. They show that this was responsible for making some of the male offspring bred from a cross between the two species sterile.

PLANT BIOLOGY

Plants shut out invaders

Cell 126, 969–980 (2006)

Researchers in the United States report that plants batten down their hatches to protect against invaders, by closing the small pores on their leaves that are known as stomata (pictured below).

Previously it was thought that stomata were simple portholes, which open to allow carbon dioxide uptake and close to prevent excessive water loss. But a team led by Sheng Yang He of Michigan State University in East Lansing has shown that the kidney-shaped guard cells around the pores on *Arabidopsis* leaves use receptors that detect bacterial compounds and trigger the stomata to close.

The researchers also show that, to get around this defence mechanism, the

plant pathogen *Pseudomonas syringae* manufactures a compound that forces the stomata to re-open after 3 hours.

VIROLOGY

Deadly response to bird flu

Nature Med. doi:10.1038/nm1477 (2006)

A comparison of patients infected by the avian influenza virus H5N1 with those suffering from human flu adds weight to arguments that H5N1's deadliness is linked to the strength of the immune response that it provokes.

Menno de Jong of the Hospital for Tropical Diseases in Ho Chi Minh City, Vietnam, and his colleagues characterized 18 cases of H5N1 and 8 of regular flu. Those patients killed by H5N1 showed the highest loads of the virus. The researchers also found elevated levels of inflammatory molecules known as chemokines and cytokines in H5N1-infected individuals, consistent with earlier results from *in vitro* and animal studies. The inflammation response may contribute to respiratory failure.

MATERIALS SCIENCE

Hydrogen's happy medium

Angew. Chem. Int. Edn doi:10.1002/anie.200601991 (2006)

Fuel tanks for the hydrogen-powered cars of the future could be made from sponge-like materials known as metal organic frameworks. A new study reveals how tweaking the materials' pore size might maximize the amount of hydrogen they can store.

Although larger pores would in theory accommodate more hydrogen, researchers



J. BURGESS/SPL

at the University of Nottingham and the University of Newcastle upon Tyne, UK, found that bigger pores have a smaller affinity for the gas. They conclude that the optimum pore size is obtained when increasing capacity is balanced with decreasing affinity. In the materials the team studied, the optimum pores had a diameter of about 0.7 nanometres.

NEUROSCIENCE

Guided trips in the brain

Nature Neurosci. doi:10.1038/nn1764 (2006)

The thin layer of tissue that surrounds the brain, known as the meninges, has an unexpected role in brain development, say Victor Borrell and Oscar Marín of the Institute of Neuroscience in Alicante, Spain.

Their discovery comes from a study of Cajal-Retzius cells, a transient population of cells that sits at the edge of the brain and guides the development of neuronal layers in the cerebral cortex. The authors tracked the migration of Cajal-Retzius cells in mice, and found that the cells move to their positions by travelling along the meninges. CXCL12, a cell-signalling molecule secreted by the meninges, seems to play a key part.

ASTRONOMY

When the giant has passed

Science **313**, 1413–1416 (2006)

Earth-like planets could exist in as many as a third of the planetary systems that have been found to harbour giant planets, new simulations suggest, despite the havoc wrought when the giants wander.

In many of these systems, giant planets that formed in the outer reaches appear to have migrated inwards, forming 'hot Jupiters'. Researchers thought such migrations were likely either to disrupt the formation of terrestrial planets in the star's neither-too-hot-nor-too-cold 'habitable zone', or to scatter such planets into unsuitable orbits.

Now researchers in the United States report simulations that suggest that Earth-like planets can form in the habitable zones after the giant has passed. Such planets, the model predicts, would be peculiarly water-rich and iron-poor compared with Earth.

CELL BIOLOGY

A twist in the tail

Cell **126**, 905–916 (2006)

Researchers have discovered a novel type of histone modification. This structural change to the proteins that package DNA can affect gene regulation.

Tony Kouzarides and his team at the University of Cambridge, UK, studied an enzyme called Fpr4 in yeast. They found that it switches the amino acid proline, found in the protruding 'tail' of histone H3, between two different mirror-image forms known as *cis* and *trans*.

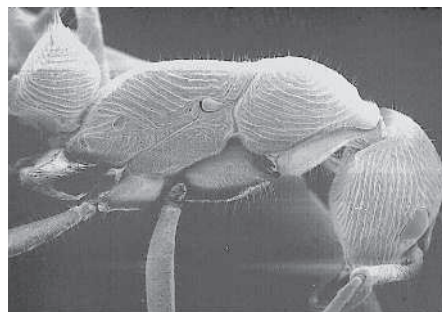
The *cis* form bends the tail sharply and allows another enzyme, Set2, to chemically alter a neighbouring amino acid, which affects the activity of genes in the surrounding DNA. The team suggests that these tail changes could also alter the structure of chromatin, the larger-scale packaging of DNA and proteins, which is known to control gene behaviour.

ANIMAL BEHAVIOUR

Once bitten

Anim. Behav. **72**, 305–311 (2006)

In queenless ant colonies, one or a few workers nevertheless corner the mating and egg-laying market. In most *Diacamma* species (pictured below), they do so by mutilating every other ant. They bite off an appendage necessary for mating soon after the ant emerges from the cocoon.



Christian Peeters of the Pierre and Marie Curie University in Paris, France, and his colleagues wondered whether the newborns fight back. It turns out they don't when the attacking ant has reproduced and might therefore be their mother, but do when the attacker is young and so probably a sibling.

The researchers reason that this makes genetic sense. The mutilated ant would only have passed on half its genes to any of its offspring, the same fraction that it shares with the offspring of its mother. But a sibling's offspring share only around a third of the ant's genes, so the victim would do better by reproducing itself.

Correction

In the Research Highlight "Changing the code" (*Nature* **442**, 960; 2006), DNA should have been described as a template that is transcribed, not translated, into RNA.

JOURNAL CLUB

Henrik Stapelfeldt
Aarhus University, Denmark

A chemist tries to see things from the molecule's perspective.

One research goal in my laboratory is to watch, in real time, how atoms exchange and electrons rearrange during a chemical reaction.

A promising technique is the femtosecond version of photoelectron spectroscopy. In the traditional version, ultraviolet light knocks 'photoelectrons' from the target molecule, revealing the molecular electronic structure.

A few years ago, Albert Stolow of the National Research Council Canada in Ottawa and his colleagues married this method with a laser whose pulses last less than 100 femtoseconds, or 10^{-13} seconds. This provided the time resolution needed to follow the flow of charge and energy during reactions (V. Blanchet *et al. Nature* **401**, 52–54; 1999).

One key element was missing, however. Scientists want to observe these processes from the molecule's frame, rather than from the laboratory. This requires that we measure the electrons' emission directions with respect to the parent molecule, a fundamental experimental problem because the molecule tends to orient randomly.

Recently, Stolow's group found a way around this problem for one process: the photoinduced cleavage of the nitric oxide dimer (O. Gefner *et al. Science* **311**, 219–222; 2006). In this reaction, the molecule breaks up, which allows its initial orientation to be determined by detection of the fragments. Combining these data with photoelectron measurements gave a picture of the electron clouds in the molecular frame — a most important milestone.

My lab recently developed a complementary solution, using laser-based techniques to hold molecules in fixed orientations. Either approach, or both combined, should get us closer to understanding reaction dynamics from the most natural point of view: the molecule's!

NEWS

Neutrinos make a splash in Italy

A long-awaited beam of neutrinos has finally made it from Switzerland to a laboratory 730 kilometres away in central Italy. The scientists involved hope the project, first sketched out 25 years ago, will address one of the big unsolved problems in particle physics. But it may yet be upstaged by the next generation of neutrino experiments.

The project, which was switched on this week, is a collaboration between the CERN particle-physics lab, near Geneva, and the Gran Sasso National Laboratory, located 1,400 metres underground in the region of Abruzzo and run by Italy's National Institute for Nuclear Physics (INFN). A flux of neutrinos — fundamental particles that are electrically neutral — is generated by an accelerator at CERN, and beamed towards the Italian lab. Neutrinos hardly interact with matter, so they travel underground through the rock in a straight line until they hit OPERA, a 1,800-tonne detector comprising emulsion films and lead plates that will stop a handful of them.

Neutrinos are known to exist in three types, or 'flavours': muon, tau and electron. Since the late 1960s, researchers studying neutrinos reaching Earth from space have noticed that they seem to change flavour, or 'oscillate' as they travel. That is surprising because when their existence was first postulated in 1930, neutrinos were thought to have no mass; and quantum mechanics says the oscillations observed require some mass, albeit a small amount.

If neutrinos do have mass, tweaking today's standard model of the Universe to take account of it could help physicists get closer to a unified theory of fundamental forces. A neutrino with mass could also explain, at least in part, why the visible matter in the Universe forms such a tiny proportion of the total mass it is known to contain.

One way to settle the matter is to create neutrinos and fire them at distant detectors. By comparing the beam's composition at source and detector, it is possible to find out whether neutrinos oscillate along the way.

The first experiment of this kind was K2K in Japan (see 'A very lengthy business'). Between 1999 and 2004, it collected evidence that muon neutrinos — the easiest flavour to study — were disappearing between the source and final destination. A similar but more powerful experiment, MINOS, started in 2005 between Fermilab in Chicago and Soudan in Minnesota. In less than a year it saw three times as



Massive project: the CERN particle-physics lab will beam neutrinos to Italy to see if they have mass.

many events as K2K, and confirmed that muon neutrinos were disappearing.

OPERA is the third and most ambitious of these experiments: it can detect not just the absence of muon neutrinos, but the appearance of taus. "The others can only see if some neutrinos are missing and presume they have changed flavour," says INFN president, Roberto Petronzio. "We can see for the first time if new neutrinos appear." Next year, ICARUS, a more sensitive detector based on liquid argon, will join OPERA. Together, it is hoped the two experiments will provide definitive evidence for muon-to-tau oscillation, and determine the tau's mass.

"We can see for the first time if new neutrinos appear."

Although the oscillation model is now widely accepted, there are still big questions. First, researchers need to rank the three flavours in order of mass. Second, they want to see whether there is a fourth kind of neutrino, which doesn't interact with matter at all. Although this type of neutrino could not be observed directly, combining results from different experiments might reveal its existence.

Great reception

For the Italian physics community, the arrival of the first neutrinos satisfies a long-held dream. As long ago as the early 1980s, Petronzio recalls, "the INFN president drew a sketch for Italian senators showing a map of Italy and Switzerland, and a line connecting

CERN and Gran Sasso". All the rooms in Gran Sasso are built so that they point to CERN.

Work on the neutrino beam got serious in the late 1990s, but funding problems slowed progress. CERN and the INFN were originally supposed to contribute equally. However, "in 2002, we decided not to have any scientific involvement in the project", recalls Jos Engelen, CERN's chief scientific officer. CERN now acts only as 'particle provider', a role he says the lab is still happy with.

Now it is finally running, OPERA may end up squeezed between old and new generations of neutrino experiments. In a few years, two new projects, NOvA in the United States and T2K in Japan, are due to start searching for the third flavour, the electron neutrino. "That is really the missing part of the picture," says Gary Barker, a physicist at the University of Warwick, UK, who is a member of the T2K group. "OPERA is a valuable experiment, but it will prove directly something MINOS has already shown indirectly."

Engelen says he is confident that the Italian experiment will provide precious results, but admits that CERN's long-term commitment is not yet confirmed. "CERN will provide the beam for the next five years, then a new agreement will be made depending on the results, both from OPERA and from the other experiments," he says. "Neutrino physics might not be among our priorities in the future."

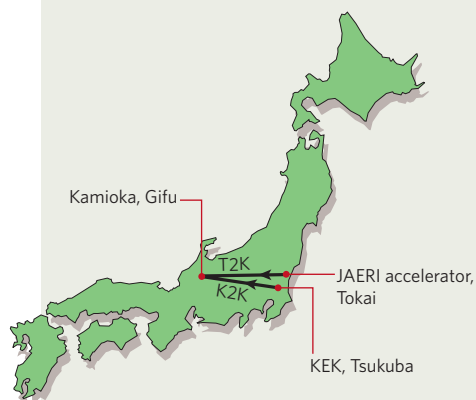
Nicola Nosengo



'PERSONAL SPACE' SURVIVES IN VIRTUAL REALITY

Real social rules are mirrored in 'Second Life'.
www.nature.com/news

A very lengthy business



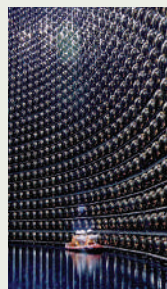
JAPAN

K2K 250 km

The first 'long-baseline' experiment, running from 1999 to 2005. Designed to confirm the neutrino oscillations reported by the Super-Kamiokande (pictured) experiment studying atmospheric neutrinos. By comparing neutrino events at the source and at the far detector, scientists could spot the disappearance of muon neutrinos, and infer their oscillation.

T2K 295 km

Due to begin in 2009, this experiment will look for electron neutrinos in a beam that initially consists of muon neutrinos after it has travelled 295 km underground. The beam is more intense than existing ones, which should maximize the probability of oscillations occurring.



SUPER-K DETECTOR

Warning flag for ethics boards

You're sitting on a university's ethics panel. A proposed experiment into the neurobiology of social behaviour comes up for approval. Subjects' brains will be scanned while they perform a series of tasks. Some of the participants have a history of disturbed social behaviour; one task involves viewing violent images. On completion, participants are simply asked to leave the laboratory. Would you give the experiment a green light?

If you're not sure, don't worry: neither are many real ethical review boards. When Canadian researchers asked their country's boards to rule on the protocol described above, they found a startling lack of conformity. All the boards that responded seemed to use similar ethical norms when judging the project, but thirty rejected it, ten approved it with qualifications, and three waved it through unconditionally (J. de Champlain and J. Patenaude *J. Med. Ethics* 32, 530-534; 2006).

"Everyone knows institutional review boards make very different decisions," says Joan Sieber, a psychologist at California State University, East Bay, who helps to train board members. She argues that it would be unrealistic to expect all boards to reach the same decision, but that the range of answers given shows that ethical norms are being applied in worryingly different ways. "This shows that something needs to be done."

The Canadian study is the work of Johane de Champlain and Johane Patenaude from the University of Sherbrooke in Quebec. The pair say they wanted to examine the belief that "considerable divergence" exists between review boards. In their experiment, each board knew that the protocol it was viewing was fictional, but was told to handle the proposal as it would any other.

Some boards worried that the consent form did not tell participants they would be viewing violent images. Another complained that the rationale for the study — to gather neurobiological knowledge that would inform "the direction in which we wish to see society develop" — reminded them of the "atrocious eugenics movement in the late nineteenth century". Two boards also noted that the rationale was not made clear in the consent forms.

Sieber says review boards need to be more rigorous in their risk assessments and decisions, and says they should call in outside experts where necessary. She adds, however, that a more common problem with institutional review boards is not a willingness to allow dubious studies, but an overly cautious approach that comes from ignorance of the experimental methods involved. ■

Jim Giles



UNITED STATES

MINOS 720 km

A beam three times longer than K2K, with higher intensity and energy, enables MINOS (pictured) to detect more events. It generates neutrinos from protons accelerated at Fermilab. The experiment looks for the disappearance of muon neutrinos, and in March 2006 it reported significant evidence of their oscillation.



FERMILAB

VLBN >2,000 km

Proposed by Brookhaven National Laboratory and at a preliminary study stage, the VLBN project aims to use a beam more than 2,000 km long (potential routes shown). It might even detect the same particles oscillating more than once along their route.

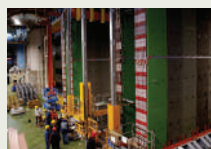
NOvA 720 km

Scheduled to begin operation in 2011, NOvA will use the same beamline as MINOS, but will have its detectors located off the beamline axis. This yields a narrow-band beam with less background noise, increasing the chance of detecting muon-to-electron oscillation.

EUROPE

CNGS 730 km

Two experiments, OPERA (pictured) and ICARUS, will analyse the neutrino beam coming from CERN using different detectors that for the first time will be able to detect the appearance of tau neutrinos, and not only muon neutrinos disappearing.



LNGS



Journey to the birth of the Universe

SOURCE: R. MCMAHON

This week, astronomers announce the discovery of the most distant galaxy yet seen from Earth. In doing so, they provide fresh clues about the early history of the Universe and seize a hotly contested record. But for some in the field, the competition to bag the record for the most distant object is losing its sparkle.

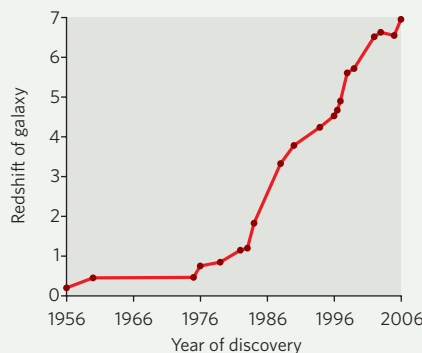
The new galaxy, dubbed IOK-1, is so far away that the light by which it was detected was emitted only 750 million years after the Universe was born (see page 186). This is just a few hundred million years after the first stars are thought to have switched on.

Other teams have claimed galaxies even farther back in time, although these are awaiting confirmation — or have turned out to be wrong (see 'The galaxy that wasn't'). But the latest detection is not in doubt. "Most people would agree that this is the most convincing, most distant object that has been published," says Richard Ellis of the California Institute of Technology in Pasadena, who also hunts for distant galaxies.

To find the object, Masanori Iye of the National Astronomical Observatory in Tokyo and his colleagues used the 8.2-metre Subaru Telescope in Hawaii to look for the characteristic radiation emitted by hydrogen in hot,

RECORD BREAKERS

Since the 1950s, astronomers have been pushing the limits of how far away we can see.



star-forming regions. This emission, known as Lyman- α radiation, has a wavelength that puts it in the ultraviolet part of the spectrum. But light reaching us from distant galaxies has been stretched, or 'redshifted', by the expansion of the Universe, so Iye's team looked for the line at longer wavelengths — the larger the redshift, the older the emitting object. They identified a Lyman- α emitter with a redshift of 6.96, then confirmed it was a galaxy by studying the spectrum of its light in more detail. The identity of



a second emitter remains uncertain.

The redshift of the most distant galaxies known has crept up from a neighbourly 1–2 in the early 1980s to the latest value obtained by Iye (see graph). Each new record has drawn much interest. "This is one of the few areas in observational astronomy where you can get your name in the *Guinness Book of Records*," notes Richard McMahon of the University of Cambridge, UK, who used to hold the record himself.

Biomedical research gets a fresh twist down on the farm

ASHBURN, VIRGINIA

Designers overlooked one small flaw when they drew up plans for Janelia Farm, the new \$500-million facility for the Howard Hughes Medical Institute in Virginia. Deer, unaware of their new neighbour built into the side of a hill, started leaping over the structure to their deaths. The deer soon wised up. But Janelia's elite group of investigators may hope they aren't taking a comparable leap into the unknown by joining this unusual laboratory.

Janelia Farm, which welcomed its first few investigators in August, trumpets itself as a rare institutional model for innovative biomedical research. The lure, including smart labs, has drawn

16 leading scientists so far, most in neurobiology and image processing, with 44 expected by 2009. Many surrendered the comforts of tenure for this cutting-edge facility, which pushes scientists to think big and take chances on interdisciplinary collaborations.

But there are risks. Investigators will be reviewed after six years initially, and subsequently every five years. If a panel of outside experts finds the research to be too conventional or progress insufficient, the investigator will be kicked out — albeit with two years of additional funding to soften the blow. "The deal is: you bet your career," says director Gerry Rubin, "and I bet \$10 million."

Those who have come to Janelia are an idiosyncratic bunch. They include Gene Myers, a shotgun gene-sequencing pioneer who gave up tenure at the University of California, Berkeley, to come. Eric Betzig started as a physicist at New Jersey's famed Bell Laboratories before working at his family's machine-tool business for eight years. Sean Eddy, a computational biologist, arrived from a tenured genetics position at Washington University in St Louis, Missouri.

By any metric, Janelia is an impressive lab space. The building is carved into a hill about

60 kilometres west of Washington DC, with each floor in the large, multi-level terrace being at ground level. Labs are walled in with thick glass from floor to ceiling, and a modular design allows them to be

easily reconfigured to suit an investigator's needs. The campus also boasts a gym, 50 apartments and a daycare facility. There's also a pub, complete with

free coffee to entice people together and a whiteboard for spontaneous brainstorming.

It's all part of a social-engineering project, painstakingly designed to stimulate social interactions and spark informal research discussions. Rubin, a *Drosophila*

"The deal is: you bet your career, and I bet \$10 million."



SOUTH KOREANS RACE FOR SPACE

Fast runners qualify for astronaut selection
www.nature.com/news

ESA/NASA/J.-P. KNEIB/R. ELLIS



Turning back time: the search for distant galaxies is giving astronomers a clearer picture of how the early Universe behaved.

The galaxy that wasn't

In March 2004, the European Southern Observatory issued a press release with the headline "VLT Smashes the Record of the Farthest Known Galaxy".

A team of astronomers had used the observatory's Very Large Telescope in Chile to study objects whose light was being amplified by the Abell 1835 galactic cluster. The

researchers thought they had found a galaxy at redshift 10, or just a few hundred million years after the Universe was born (R. Pelló *et al. Astron. Astrophys.* **416**, L35–L40; 2004). But later analyses of the data cast doubt on this claim, and recently the Spitzer Space Telescope failed to see the object at all.

According to first author

Roser Pelló, the team was confident enough to publish because three lines of argument supported its conclusion. Now she says that they don't know what they saw: "This is no longer a redshift 10 identification. It does not mean the source is spurious, it just means it is a puzzling, faint object." **J.H.**

by Rychard Bouwens and Garth Illingworth of the University of California, Santa Cruz (see page 189). They analysed data from the Hubble Space Telescope and suggest that the number of detectable galaxies drops off above a redshift of 7.

This decline might mean that astronomers are getting close to the point when the first galaxies formed. Earlier galaxies may have been smaller, fainter objects that later merged to form brighter beacons. Such information could be fed into simulations of how hydrogen in the early Universe clumped together (see page 151). "What we've done is provide key data that can provide the constraints or inputs to the models," says Illingworth.

That doesn't mean astronomers have given up on going farther back in time. New instruments, including an infrared camera to be installed on the Hubble Space Telescope — if the space shuttle makes another servicing trip — and the James Webb Space Telescope, due to fly in 2013,

should find a flood of more distant objects.

In the meantime, new techniques are pushing back the limits. Ellis, for example, is searching for distant galaxies whose faint light has been amplified by the gravitational fields of closer galactic clusters. He is preparing a paper that claims six candidates around a redshift of 10 — dating from when the Universe was between 400 million and 500 million years old.

Some prefer to investigate closer, brighter galaxies, about which more can be learnt. "Unless you can measure something else in the galaxy, it's not very useful," says Illingworth.

But going for the record does get you attention. A story featuring Ellis's search for the oldest objects made the cover of *Time* magazine's 4 September issue. Illingworth, an old friend and collaborator, jokes: "It's good for astronomy to have these articles, but I'll kick Richard in the butt sometime for not mentioning the other things we're doing."

Jenny Hogan

But some hope to move on from merely claiming records. "I think the field is changing from being simply record-driven, to finding out whether we can reach times where we can see very different properties in the early Universe," says Esther Hu, of the University of Hawaii in Honolulu, who championed early efforts to find Lyman- α emitters.

In this respect, the more interesting aspect of Iye's paper is that the team found fewer galaxies than it expected. This matches a survey

geneticist who is moving his lab to Janelia from the University of California, Berkeley, based Janelia loosely on the open laboratories of Bell Labs and the Medical Research Council's Laboratory of Molecular Biology in Cambridge, UK.

Groups are no bigger than six members, to encourage interdisciplinary collaboration that relies on field experts rather than graduate students or postdocs. This should also ensure that group leaders are active bench scientists

rather than just managers, says Rubin. The glass walls help too, by letting researchers see who's in and what they're doing.

At Bio-X, a similar suite of glass-walled labs at Stanford University in California, few have complained about the lack of privacy, says its operations director Heideh Fattaey. But some have set up whiteboards to block off space. "You can't change human nature," she says. In-house architect Bob McGhee, who helped to design the labs at Janelia, acknowledges that they are not suitable for researchers conducting, for example, ethically sensitive animal studies.

Rubin insists that he doesn't want to replicate other institutes, no matter how successful or prestigious they may be. "My greatest failure would be if

this turns into another Salk or Whitehead," he says.

Interdisciplinary labs may be trendy, but placing physicists next to biologists doesn't necessarily spark a revolution. "There's not as much real collaboration as one would think," says Janelia group leader Loren Looger, a protein-engineering researcher with a background in maths and chemistry, who has also worked at Bio-X. But Janelia should prove more successful, he says, as group leaders, already proficient in multiple disciplines, will not have to worry about defending unconventional research to secure tenure — sometimes a problem in academia.

"It may fail," says Looger. Nevertheless, he revels in being part of the Janelia experiment. **Gene Russo**

P. FETTERS



Space to think: the Janelia Farm lab aims to encourage unconventional work.


AMERICAN CHEMICAL SOCIETY

Catch up on the meeting with our daily diary reports. <http://blogs.nature.com/news>

F. CHMURA/ALAMY

SNAPSHOT
Women at work

Scenes such as this one, of the opening ceremony of a Dutch academic year, might become more common if the 'Lisbon target' of 25% female faculty in European universities is ever achieved. European research ministers set this 2010 goal, now looking unlikely, in June 2000.

But this particular ceremony is staged. Petra Rudolf, a professor and materials scientist at the University of Groningen in the Netherlands, chivvied 35 of the university's 50 female professors into showing up at the opening ceremony of the new academic year on 4 September. Rudolf's whip pushed the gender ratio of those in gowns to more than a quarter.

"The men seemed, shall we say, to feel the difference," laughs Rudolf, "and they were not entirely comfortable." But they'll have some time to get used to it. The university expects to increase female faculty from 10%, the Dutch average, to 15% by 2010.

Alison Abbott



R. BOXEM/HOLLANDE HOOGTE

Extreme TB strain threatens HIV victims worldwide

An outbreak of virulent and very drug-resistant tuberculosis in South Africans with HIV is horrifying health professionals and prompting the usually slow-and-steady field of tuberculosis medicine into emergency action. "This is the nightmare scenario of HIV-fuelled drug resistance," says Paul Nunn of the World Health Organization's (WHO's) Stop TB Partnership.

The new strain, called extreme drug-resistant tuberculosis (XDR TB), is virtually untreatable. Neither the standard drugs nor at least three of the six classes of more toxic and less-effective backup drugs are effective.

So far all those infected in the South African outbreak have HIV. Drug resistance is energetically costly for bacteria to maintain, and it may be that a strain resistant to so many drugs is not virulent enough to make healthy people seriously ill. But those with immune systems weakened by HIV are more susceptible to bacterial infections, and in populations with high concentrations of HIV, patients infected with XDR TB will pass it on to those around them.

Delegates to an emergency conference in

Johannesburg heard last week that more than 120 patients in 28 hospitals in rural Kwazulu Natal province, South Africa, had contracted this severe form of tuberculosis. Nearly all of those known to be infected are now dead, but many others may be carrying the bacteria. "There is no systematic survey, so this is just the tip of the iceberg," says Nunn.

A survey released in March showed that XDR TB is a global phenomenon, with relatively more cases occurring in Eastern Europe and Western Asia. But until now, only isolated cases of XDR TB have been seen. The Kwazulu situation is the first outbreak of XDR TB. It was first reported at last month's International AIDS conference in Toronto, Canada, but the latest figures, from a wider area, give a clearer picture of the scope of the outbreak. The fear is that the disease could spread north where HIV levels are also high, if it hasn't already.

Scientists at last week's emergency conference, organized by the South African Medical Research Council, the WHO and the US Centers for Disease Control and Prevention,

struggled to gear up efforts fast enough to tackle the outbreak.

"The TB community is used to doing things slowly and steadily, and thoroughly and well, not at all in outbreak mode such as with SARS and avian flu, and I think that is required," says Nunn. But he says it is hard to galvanize efforts in Kwazulu because of the problems that are typical across the developing world — including overcrowded hospitals, unfinished treatment courses, lack of well-equipped labs, widespread HIV infection and grinding poverty. "There is a certain amount of fatigue."

Officials at the nonprofit Global Alliance for TB Drug Development in New York say that they have been approached to see if any of the drugs they are researching could be deployed on a 'compassionate use' emergency basis.

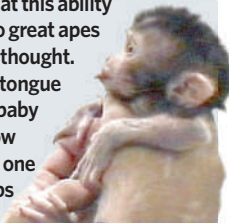
Unfortunately, according to alliance president Maria Freire: "The drugs we have are too early for that." The alliance estimates that the first drugs in an entirely new class will not be ready for regular use until 2012.

Emma Maris

ZOO NEWS

Making faces

Italian researchers have found that infant rhesus macaques can imitate facial movements, suggesting that this ability isn't unique to great apes as previously thought. Sticking your tongue out all day at baby monkeys? Now that has to be one of the best jobs in the world.



Live fast, die young

The most distant galaxy isn't the only scientific record notched up this week (see page 128). Guinness World Records has confirmed that the coral reef pygmy goby (*Eviota sigillata*), discovered by Australian researchers, is the vertebrate with the shortest known lifespan — a fleeting 59 days.

ON THE RECORD

“Experience your own choice of a dazzling and varied array of gourmet food.”

Invitation for the official party of the 10th International Congress on Obesity in Sydney.

“To vomit in space is not my idea of a good time.”

Star Trek actor William Shatner turns down British entrepreneur Richard Branson's offer of a real space flight.

OVERHYPED

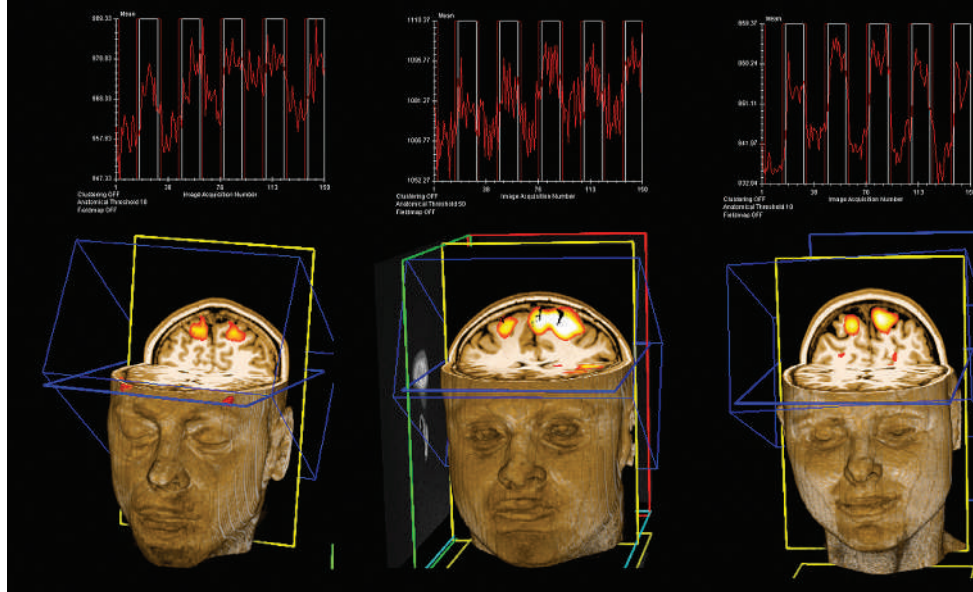
Last week, news sources worldwide reported that a parapsychologist at Trinity College, Cambridge, had proved the effect of telephone telepathy — in which a subject ‘knows’ who is calling before they answer the phone.

But the researcher, Rupert Sheldrake, is not employed by the university. Instead, he gets a grant from the Perrott-Warrick fund for psychic research, set up at Trinity about a century ago. The finances are still looked after by the college, but it has no connection with the committee that spends the money.

Martin Rees, master of Trinity, says it is inappropriate to describe Sheldrake's affiliation as Trinity College. “I think he's a former fellow of Clare, which should have undiluted credit,” he says.

Sources: PLoS Biol., UPI

“Imagine playing tennis....”



Brain areas activated in healthy volunteers (above) were identical to those in a vegetative patient.

‘Vegetative’ patient shows signs of conscious thought

The question of whether outwardly unresponsive patients may in fact be aware of their surroundings is one of the most heated debates in clinical neuroscience, with huge implications for the way such patients are diagnosed and treated. In a remarkable study published last week, researchers report that a patient who meets all the criteria of being in a ‘vegetative state’ can perform mental tasks on request. But the case highlights the difficulty of probing the mental state of an individual who cannot communicate, and of drawing any general conclusions about his or her condition.

Adrian Owen of the Medical Research Council's Cognition and Brain Sciences Unit in Cambridge, UK, and his colleagues used functional magnetic resonance imaging (fMRI) to scan the brain of a 23-year-old woman left in a vegetative state after a car accident in July 2005. Such patients do not respond to their surroundings, and doctors have always assumed that they are completely unaware of them. Previous brain-scanning studies have failed to detect more than reflex reactions to stimuli such as sound or pain.

Owen and his colleagues now show that their patient can respond to verbal requests to

picture herself either playing tennis or walking around her house. Each activity produces a different pattern of brain activation, and these are strikingly similar to those of healthy volunteers (A. M. Owen *et al. Science* **313**, 1402; 2006). The decision to comply with the experimenter's instructions proves awareness, says Owen. “What we've developed is a method for detecting when someone is aware in the absence of other clinical evidence.”

But what that ‘awareness’ means is still up for debate. For example, Paul Matthews, a clinical neuroscientist at Imperial College London, argues that the brain imaging technique used cannot evaluate conscious thought; fMRI lights up regions of brain activity by identifying hotspots of oxygen consumption by neurons. “It helps us identify regions associated with a task, but not which regions are necessary or sufficient for performing that task,” he says.

Matthews argues that the patient's brain could have been responding automatically to the word ‘tennis’, rather than consciously imagining a game. He also points out that in many vegetative cases, the patient's motor system seems to be undamaged, so he questions why, if they are conscious, they do not respond

A.M. OWEN

physically. "They are simply not behaving as if they are conscious," he says.

Owens counters that an automatic response would be transient, lasting for perhaps a few seconds before fading. He says his patient's responses lasted for up to 30 seconds, until he asked her to stop. He believes this demonstrates strong motivation.

He does admit, however, that it is impossible to say whether the patient is fully conscious. Although in theory it might be possible to ask simple 'yes/no' questions using the technique, he says: "We just don't know what she's capable of. We can't get inside her head and see what the quality of her experience is like."

If she is conscious, a pressing question is how many other vegetative patients might be in a similar condition. Here the uncertainty is even greater. A vegetative state can be triggered by a vast range of brain injuries, which vary widely in severity.

"Family members should not think that any patient in a vegetative state is necessarily conscious and can play tennis," says co-author Steven Laureys of the University of Liège, Belgium. "It's an illustration of how the evaluation of consciousness, which is a subjective and personal thing, is very tricky, especially with someone who cannot communicate."

The patient studied by Owen's group, for example, seems to have been much less severely injured than the permanently vegetative Terri Schiavo, whose case inflamed the moral debate over whether those who show no signs of recovery should be allowed to die. Schiavo's feeding tube was removed in March 2005 after a legal wrangle.

Nevertheless, the research raises questions about whether the definition of a vegetative state should be changed to allow the possibility of using brain imaging to ascertain awareness. Currently, the condition is diagnosed if patients show no external awareness of themselves or their environment, although, in contrast to coma patients, their sleep-wake cycle is preserved. Typically, after a month the patient is judged to be 'persistently vegetative', and the condition is generally pronounced 'permanent' after a year without signs of recovery. The British and US legal systems then allow doctors to question whether the patient's treatment should be continued.

The British patient in the latest research seems to have been at the less severe end of the range

of injuries, and is now showing signs that she will recover, says Laureys. Since the research was carried out at the end of 2005, she has developed the ability to track her reflection in a mirror — which suggests she is emerging into a 'minimally conscious state'. Laureys admits it is therefore possible that her brain scans merely heralded the early stage of her recovery, rather than providing real insight into the vegetative state. "We cannot exclude that she was in transition," he says. Another possibility is that the contact made during the study itself might have helped trigger the improvement — a factor that would raise the hope of being able to treat some cases. But that "remains pure speculation", says Laureys.

Either way, the case is likely to serve as a reminder to doctors that the vegetative state of patients should not be taken for granted. In 2003, another patient, Terry Wallis, stunned doctors by making a marked recovery after spending 19 years in a minimally conscious state. Neuroscientists later discovered that the neurons in his brain had been slowly reconnecting over the years, to bypass damaged regions (H. U. Voss *et al. J. Clin. Invest.* **116**, 2005–2011; 2005).

Laureys believes that many patients may be slowly moving along a continuum from the vegetative condition, through minimal consciousness to full recovery. "As clinicians, we're obliged to reduce reality to a binary condition," he says. "We have to draw a line where there is actually a grey zone."

Neuroscientists have been trying for years to find firm evidence about awareness in vegetative patients. But

previous studies using positron emission tomography have proved inconclusive. Four years ago, Laureys reported that although primitive brain centres in such patients were activated in response to a painful electric shock, the frontoparietal network, thought to be required for conscious perception, was not. This suggested that patients were unaware of the pain to which they had responded (S. Laureys *et al. NeuroImage* **17**, 732–741; 2002).

Studies using fMRI, which is thought to give a more accurate picture of brain activity, have been developed only recently. Owen says he hopes that real-time fMRI, which has become available since his latest study, will allow more sophisticated tests. ■

Michael Hopkin

See Editorial, page 121



"What we've developed is a method for detecting when someone is aware in the absence of other clinical evidence."

— Adrian Owen

Autism comes under three-pronged attack

The possibility that an antibiotic could help treat autism is the basis for one of three new US clinical studies of this brain disorder.

Minocycline is being considered as a potential treatment for autism because of its anti-inflammatory powers. Although no single cause of the condition has been identified, it has been linked to immune-system activity that causes inflammation in the brain. Researchers at the National Institute of Mental Health (NIMH) in Bethesda, Maryland, who announced the trial on 6 September, will initially test the antibiotic in 10 children. If they show improvements in areas such as communication skills, a larger, placebo-controlled study will be launched.

A second study will focus on biological and behavioural markers specific to different subtypes of autism. And in the third study, NIMH researchers will assess chelation therapy, which removes heavy metals from the blood. This is sometimes used by families of autistic children, but its effectiveness has not been studied in detail.

South Africa responds to critics of its AIDS policies

The South African government last week moved to curb the activities of its controversial health minister, Manto Tshabalala-Msimang, three days after 82 scientists worldwide called for her removal in a letter to President Thabo Mbeki. They charged that she had “undermined” HIV science by promoting nutrition-based AIDS treatments, and was thus an “embarrassment”.

The scientists were spurred to write the letter after attending a major AIDS conference in Toronto, Canada, last month, where the South African government exhibition featured garlic, lemons and African potatoes — “with the implication



Under fire: protesters want Manto Tshabalala-Msimang sacked for implying foods can treat HIV.

P. MARTINEZ MONSIVAIS/AP PHOTO

Collaboration aims knockout blow for mouse genes

A massive international project has been launched to generate at least one mutation in each of the more than 20,000 genes of the mouse. It will be the largest international research collaboration in biology since the Human Genome Project.

The project was announced on 7 September by the European Commission, the US National Institutes of Health and Genome Canada, which together will provide US\$72 million. The aim is to coordinate the production of mutant mice, like the obese mouse shown here, and make them available to researchers. Methods include knockout technology, in which specific genes are deleted, and treating mice with compounds that cause mutations at random. Some 12,000 mouse mutants have so far been generated in different labs.

The work will allow geneticists to more accurately dissect gene function in a mammal that shares 99% of its genes with humans, and so better understand the genetic basis of human disease.



that these dietary elements are alternative treatments for HIV infection”, the scientists complained. Prominent signatories included Nobel laureate David Baltimore, virologist Robert Gallo and leading South African physicians such as Hoosen Coovadia of the University of KwaZulu-Natal.

Mbeki has reportedly asked his deputy president, Phumzile Mlambo-Ngcuka, to set up a ministerial council to oversee the government's anti-AIDS efforts. Baltimore said he was “gratified” by the move.

Science superstars weigh in to sell chemistry to kids

It doesn't look like a run-of-the-mill attempt to engage the attention of youngsters. The Molecular Frontiers initiative, launched at the First European Chemistry Congress in Budapest late last month, is a global think-tank of scientific heavyweights who will monitor the frontiers of molecular — mostly chemical — sciences, and generate ways of stimulating the interest of children.

Seven of the 23 members of the scientific advisory board are Nobel prizewinners. Other notables include fractals expert Benoit Mandelbrot and genome entrepreneur Craig Venter.

Most activities will be Internet-based. Prizes of US\$1,000, for instance, will be given to children who come up with the best question about molecular science.

► <http://web.mit.edu/molecularfrontiers>

China and India agree on joint technology push

In a fresh initiative to boost cooperation in science and technology, India and China have agreed to set up a ministerial-level steering committee jointly chaired by the science ministers of the two countries.

Indian science minister Kapil Sibal and

his Chinese counterpart, Xu Guanhua, signed the agreement on 7 September in Beijing. It revives an 18-year-old pact that has previously made little progress.

A government spokesman in New Delhi says the steering committee will prepare a road map for future cooperation in science and technology before Chinese president Hu Jintao visits New Delhi in November.

India and China are expected to pool their respective strengths in computer software and hardware in the hope of emerging as a regional leader in information technology. Also under discussion is the creation of a Sino-Indian nanoscience forum, and collaborative projects in agriculture and genomics.

Gene-sifting tests get regulatory treatment

Complex diagnostic tests that sift through data on multiple proteins and genes are facing formal regulation by the US Food and Drug Administration (FDA).

The tests, known as multivariate index assays, are relatively new and are typically developed and carried out by individual laboratories. They are used for a range of conditions including Alzheimer's disease and breast cancer.

The FDA had previously said that in principle it could regulate such assays, but it wasn't until 5 September that it finally issued draft guidelines. The agency says the regulations are needed because the tests are becoming more numerous and are increasingly used in serious conditions.

“It is important for the companies and labs making the tests to clearly understand the regulatory requirements in place so that the tests they develop are as safe and effective as possible,” says Daniel Schultz, director of the agency's Center for Devices and Radiological Health. The draft guidelines are open for 90 days for public comment.

BUSINESS

Nanotech's big issue

Lack of regulation and risk assessment could hamper the nanotechnology sector. **Virginia Gewin** reports.

Nanotechnology is at last set to start delivering on its promise as a growing number of products hit the market place. But even as companies put the finishing touches to their ideas, there is growing unease that the sector is ill-prepared for the rigours of the public arena.

New technologies carry with them an attendant risk, and nanotechnology is no exception. At the start of this year, 'nano' was an advertising buzz word (see *Nature* **440**, 262; 2006) — now, that gloss is already beginning to fade.

In March, a number of German consumers experienced respiratory problems after using the glass sealant Magic Nano, and the product was pulled from the shelves. Although later tests showed that, despite its name, the product contained no nanoparticles, the incident heightened public awareness that nano-related products might carry unforeseen risks.

Mindful of the lessons learned over the marketing of genetically modified food, the nanotech sector is taking steps to address the risks, real or perceived, associated with its products. Some consumer advocates, for example, have questioned whether nano-sized particles might behave differently in the body from larger particles if ingested or inhaled. And many companies are taking such concerns seriously.

German chemical giant BASF, for example, is developing a number of nanotech applications and publishes extensive information on its website about its activities. And chemical firm DuPont has joined forces with the non-profit group Environmental Defense in New York to develop a framework for companies to assess nanotech risk.

These companies have had to act on their own initiative because there are no regulations anywhere in the world to govern nanotech-related products. This lack of regulation worries observers and investors alike.

"Nobody wants to see over-regulation, but everyone is aware that the biggest risk for nanotechnology is no regulation," says Christoph Meili, chief executive of the Innovation Society, a nanotechnology consulting group based in Switzerland. The greatest concern focuses on nano-sized particles used in consumer products, such as cosmetics or mineral supplements, that fall outside the existing purview of agencies such as the Environmen-



BASF has opted to publicize its nanotech research on the web.

tal Protection Agency (EPA) or the Food and Drug Administration (FDA).

There are some moves to address this situation. The EPA and the UK Department for Environment, Food and Rural Affairs are exploring schemes whereby companies would voluntarily report risk findings. The FDA recently assembled a task force to determine regulatory approaches to safeguard products using nanotechnology. And last summer the European Commission released a nanotech action plan to serve as a framework for crafting future regulation.

So far, health and safety issues have not hampered overall investment in nanotechnology companies, according to a report this May from analysis firm Lux Research based in New York. Although the number of nanotech-related deals declined slightly in 2005, total venture-capital investment increased from around \$410 million to almost \$500 million (see graph).

A follow-up report from Lux, published in June, highlighted the increasing worries about product safety. It quoted anonymous nanotech-

savvy executives saying that some companies had halted their nanotech plans over confusion on how to judge potential hazards. For example, one interviewee said his firm had stopped developing a product when it became clear that it would cost too much to ensure the product would be risk-free for its entire life.

Such concerns can dissuade investors. Steve Jurvetson, managing director of Draper Fisher Jurvetson, a venture-capital firm in Menlo Park, California, says his group simply doesn't invest in areas with an unspecified regulatory regime because it's not worth the risk. Like many investors tempted by the sector, he sticks to supporting nano-sized improvements to existing products rather than novel nanoparticles that, as yet, have little direct application.

Perceived risk is a formidable issue that must be addressed, says Michael Holman, an analyst at Lux. The Magic

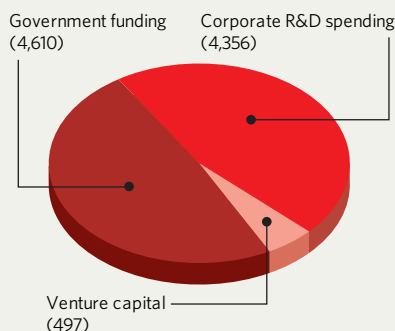
Nano incident raised the business community's awareness that the public needs to be better educated about nanotech's benefits and risks. Some note that the word 'nano' is no longer a hot marketing term. "It's not helpful any more to use nano for your company name," says Juan Sanchez, a vice-president at Punk, Ziegel and Company, an investment bank in New York.

More risk research could only help. In 2005, according to its own estimate, the US government spent some \$40 million on research into the risks of nanotechnology. But an independent review of research documents found only \$11 million that had clearly been spent on those areas, says David Rejeski, director of the Woodrow Wilson International Center for Scholars' Project on Emerging Nanotechnologies in Washington DC. His group has called for a minimum investment of \$100 million over the next two years; Dupont and Environmental Defense have called for the same amount to be spent annually.

As nanoproducts develop beyond films and coatings to more novel nanostructures, analysts suggest that more risk research is in the industry's best interest. The EPA approved its first nanoproduct, the manufacture of a specific carbon nanotube, last autumn under a clause of the Toxic Substances Control Act. Any future products will have to be subject to this act and myriad current regulations from different agencies — as well as any likely future nano-specific amendments.

But with the number of products ready for market set to increase, many argue that the sector must act quickly. "My fear is that once we figure out how to do first-generation oversight," says Rejeski, "the second generation will be here."

GLOBAL NANOTECH INVESTMENT IN 2005 (US\$ million)





NHM, LONDON

DIGGING FOR DODO

No one has seen a dodo in three and a half centuries, but that hasn't stopped the bizarre speculation about this extinct bird. **Henry Nicholls** investigates whether recent excavations in Mauritius could reveal the real creature.

The dodo is perhaps the most famous symbol of extinction, yet no one knows exactly when or how it went extinct. In fact, no one knows much of anything about the flightless creature, which lived on the Indian Ocean island of Mauritius until soon after Portuguese and Dutch sailors showed up in the sixteenth century. Popular images perpetuate the notion that the dodo (*Raphus cucullatus*) was a tubby, pigeon-like bird with a huge bill, gaping nostrils and bulging eyes.

Yet this clumsy apparition may be pure fiction, created from secondhand stories and poorly preserved specimens (see 'Dodo fiction', opposite). Thousands of dodo bones, hundreds

of dodo studies and countless dodo enthusiasts have done little to illuminate how the dodo really lived and died. "We still know virtually bugger all about the dodo," says Julian Hume, a bird palaeontologist at the Natural History Museum in London.

That may soon change, thanks to new excavations in a Mauritian swamp called the Mare aux Songes. For the second year running, a team led by Dutch geologist Kenneth Rijdsdijk has been visiting the most important site ever to turn up dodo bones. The spot has languished untouched for decades, but in the years to come, mud from this dig could increase the number of known dodo bones by an order of

magnitude. And there is an entire ecosystem, from bacteria to giant tortoises, preserved alongside the dodo remains. Together, the studies promise an unparalleled glimpse into the life and death of this enigmatic bird — perhaps even providing, for the first time, a faithful picture of the long-dead dodo.

It may seem surprising that scholars disagree about so many basic facts, including what actually drove the dodo to extinction. There are many possibilities. The newly arrived sailors may have eaten so many birds that they caused the population to crash. Or perhaps the sailors destroyed its habitat beyond repair, or brought mammals — notably rats, pigs and

goats — that trashed the landscape. Or a natural disaster, such as a tropical cyclone, may have pushed the bird to near-extinction before the sailors even arrived. The smart money is on the rats, but without the luxury of travelling back in time, it's impossible to be sure.

Opinion is also divided over the date the dodo finally disappeared¹. Hume has argued in *Nature* that detailed hunting records from a chief of Mauritius suggest his men bagged at least a dozen dodos between 1685 and 1688 (ref. 2). Others disagree, including Hume's friend and colleague Anthony Cheke, leader of a 1970s expedition to Mauritius by the British Ornithologists' Union. Cheke says that the word in the hunting records — *dodaersen* — was used by the 1660s to describe another flightless bird on the island, the red rail (*Aphanapteryx bonasia*)^{3,4}. The last reliable dodo sighting, he argues, dates from 1662, when a band of shipwrecked sailors waded out from Mauritius to a small islet and hunted down several birds "larger than geese".

Up in smoke

Whenever and however it vanished, the dodo did not leave much of itself behind. Sailors brought back stories, sketches and stuffed birds. Only one bird is known to have been brought to Europe alive; it may be this creature whose skeleton wound up in the Ashmolean Museum in Oxford, UK. Tragically, in 1755, all but the head and one foot of this precious specimen went up in flames. One account of the story involves a heroic curator, who dived into the flames and hacked off the bits, saving them from certain destruction. The more widely accepted version is that the fire was no accident, but was deliberately lit to destroy a deteriorating specimen.

The salvaged body parts remained the world's most important dodo relic for more than a century. Then in 1865, engineers were carving out a railway line alongside the Mare aux Songes on the east coast of Mauritius

(see map). Local schoolteacher George Clark chanced upon dodo bones exposed during the digs, and he organized local workers to feel their way through the nearby swamp in search of more material.

Clark eventually uncovered an impressive cache of dodo bones, which he sold to the Natural History Museum in London. It is largely this and other Mare aux Songes material extracted towards the end of the nineteenth century — a few thousand bones in total — that informs modern understanding of dodo anatomy. But because of the crude techniques used by Clark's crew, the Mare aux Songes has never produced a complete skeleton of a bird. Instead, almost all the dodo skeletons on display in the world's museums are the remains of several birds unearthed from the swamp by Clark and his immediate successors, and cobbled together into a dodo-like form.

There is one exception, and it stands inside a six-foot glass case in the Mauritius Institute in the capital city of Port Louis. This specimen, found in the 1890s by a hairdresser and dodo enthusiast named Louis Etienne Thirioux, is the only known fully articulated specimen. It probably came from a cave at Le Pouce just south of Port Louis, says Hume. But Thirioux seems to have kept the exact location a secret, and the few people who have bothered to retrace his steps have turned up nothing as the valleys are now overgrown with impenetrable weeds.

Scant remains

Others have continued to search for dodo remains elsewhere on the island. In 1974, a handful of dodo material came out of a borehole at Mare Sèche, but the landowner would not allow palaeontologists to excavate. "You can find bones in other places, but in very limited quantities," says Anwar Janoo, an ornithologist and consultant for the National Heritage Fund in Mauritius. He has found, for instance, some dodo fragments in cliffside caves at Baie du Cap at the southern tip of the island. Slaves and convicts used these caves as refuges, he says, and the bones are probably the remains of dodos captured and eaten by fugitives. Janoo has also found some bones at the bottom of a collapsed lava tunnel on Plaine des Roches in the north⁵.

He claims to know of another dodo site in the north where French palaeontologist and filmmaker Didier Dutheil exhumed the back of a dodo skull in 1999. But neither Janoo nor Dutheil are letting on precisely where it is. The secrecy is frustrating, says Cheke: "It's amazing they haven't gone back or at least told somebody else where it is."

A few thousand bones from Mare aux Songes, the Thirioux skeleton and Janoo's fragments are clearly not a lot to go on. But it wasn't the need for more specimens that led to the recent digs at the Mare aux Songes. Instead, a series of coincidences brought Rijdsdijk, who is based at the Netherlands Institute of Applied Geoscience in Utrecht, and his colleague Frans Bunnik there

DODO FICTION

Despite a wealth of popular writing about the dodo, secondhand or conflicting stories have led to plenty of misunderstandings about what the bird was really like⁷.

Colour All firsthand reports agree that the dodo was grey or blackish, but seventeenth-century Dutch paintings of white dodo-like birds led to confusion. Some of these pictures may have shown the now-extinct white ibis (*Threskiornis solitarius*) from nearby Réunion Island. Others may have portrayed an albino dodo, perhaps singled out because of its unusual colour.

Shape The popular image of the dodo — a squat, short-legged, awkward-looking beast — probably derives from one image created by Flemish painter Roelandt Savery around 1626 (opposite), says dodo expert Julian Hume. Earlier illustrations show long, powerful legs and a more upright stance. But copying soon became rife. "After 1638, you cannot trust any image of a dodo as being genuine," says Hume.

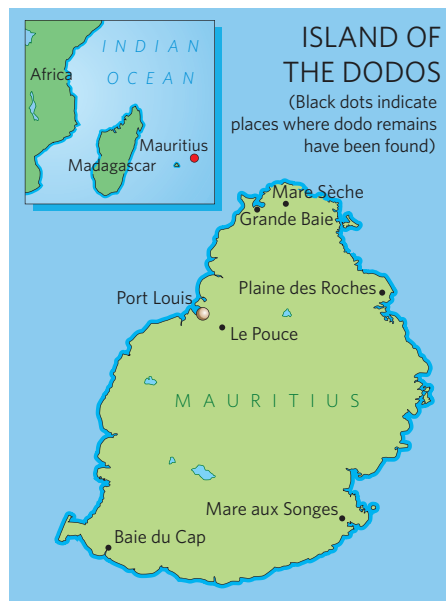
Nostrils In many illustrations, the dodo has huge, gaping nostrils. This is probably because almost all artists based their work on dead specimens, says Hume. "When the skin dries, the holes open up," he says. In the only pictures unequivocally based on live or recently dead dodos, such as that dating from 1601 below, the nostrils are almost invisible.



Diet With virtually no information on what dodos ate, speculation has been rife. One theory, proposed in 1973, held that the seeds of the tambalacoque tree needed to pass through a dodo's digestive tract before they would germinate, and without dodos, the tambalacoque was declining. The recent rediscovery of tambalacoque seedlings in Mauritius dealt a final blow to this fable.

Edibility The dodo probably didn't taste as awful as popular legend holds, says Hume. One of the earliest accounts describes the birds as having "very good meat" but goes on to say that sailors preferred to eat pigeons and parrots, as the dodos were "loathsome and fulsome birds". Hume suggests that this phrase may have been misread: dodos tasted fine, he thinks, but were so rich and fatty that sailors quickly tired of the meat.

H.N.



R. JAYASENA



R. JAYASENA

Skeleton key? Julian Hume (right) holds up one of the many dodo bones retrieved from rich soil samples (left) taken from a Mauritian swamp.

in October 2005. They were looking not for dodo, but for pollen samples that might help them reconstruct the vegetation and climate of Mauritius before the Dutch settled the island in 1638.

Rijsdijk and Bunnik identified several spots of interest, but neither of them realized that one of their sites — a marshy valley not far from the sea and surrounded by sugar-cane plantations — was the swamp Clark had excavated 140 years earlier. In the 1940s, the then British authorities ordered the Mare aux Songes to be filled with volcanic rubble, to combat the spread of mosquitoes and malaria on the island. Rijsdijk and Bunnik realized where they were when a local landowner told them that five boreholes had been drilled into the swamp in 1992, at the behest of a dodo devotee from Japan. The person who commissioned the work never claimed the cores, and the Dutch scientists were the first to see them.

They were stunned by the richness of material in the cores. “They had an alternation of different types of sediment that showed we were dealing with a really dynamic environment,” says Rijsdijk. There was one small problem: knowing where the cores had come from. A crude diagram made by the drill engineers sketched out the sea, a road, four palm trees and the boreholes — with no scale or direction. “It was like a treasure map,” he says.

Dodo hunt

Locating a likely spot, they began to drill a speculative core. “It took us one and a half hours to get through 80 centimetres of rocky rubble,” says Rijsdijk. But beneath they found dozens of dodo bones.

This past July the team returned, armed with radar to locate the most promising areas, a digger to remove the top layer of rubble, and drills to bore into the peaty soil to depths of some eight metres. The cores came back up chock

full of microbes and plant matter, along with bones from skinks, giant tortoises, parrots, owls and bats. Radiocarbon dates should soon pin down the age of these deposits, but the cores were completely free from introduced species — suggesting that they represented the dodo’s world before humans arrived. “This provides us with a great opportunity to reconstruct this world,” says Rijsdijk.

In fact, the Mare aux Songes may capture more than just a moment in the life of the dodo, says David Burney, a palaeoecologist at Fordham University in New York. “I’m predicting it’s not just a snapshot,” he says. “It may be a kind of movie, sampling the landscape more or less continuously back in time.” Burney works in Hawaii, studying how the ecosystem was changed by the arrival of the

Polynesians; similarly, the studies on Mauritius may reveal the effect that natural forces and human colonization have on island ecosystems.

This summer’s excavations focused on an area of swamp about the size of a small room. It yielded about 4,000 bones, of which a few hundred belonged

to the dodo; early results will be presented at a conference in Oxford later this month. All told, there are three basins in the Mare aux Songes with palaeontological promise; they cover about 5 hectares and could contain tens of thousands of dodo relics, says Rijsdijk.

Such bounty could plug some of the holes in scientists’ understanding. Were male and female dodos different sizes? Did males sport an exaggerated bill or another trait to attract partners? How big were the dodos’ eggs? And how did the Mare aux Songes end up with so much interesting material — was it a natural disaster? If so that would lend credence to the idea that dodos had been battered to near-extinction before humans arrived.

Some hope that the swamp could even yield dodo DNA. An earlier study of dodo genetics, with DNA taken from the foot of the Oxford

dodo, illustrates the potential of this research. It suggested that the dodo’s closest living ancestor is the Nicobar pigeon (*Caloenas nicobarica*), and that the two diverged more than 30 million years ago⁶. But this date is long before volcanic Mauritius emerged from the sea. New studies could help clarify this relationship, or even yield new information, says Beth Shapiro, a specialist in ancient DNA with the University of Oxford. “If we were able to get ancient DNA from a lot of dodos from a long time period, we might be able to see how the population size changed as it approached extinction,” she says.

But the hot climate and acidic conditions of the Mare aux Songes will not have helped DNA preservation. Early analysis of the mud yielded DNA sequences from several different plant species, but so far nothing that looks as if it could have come from a dodo. “We have not yet given up,” says Shapiro.

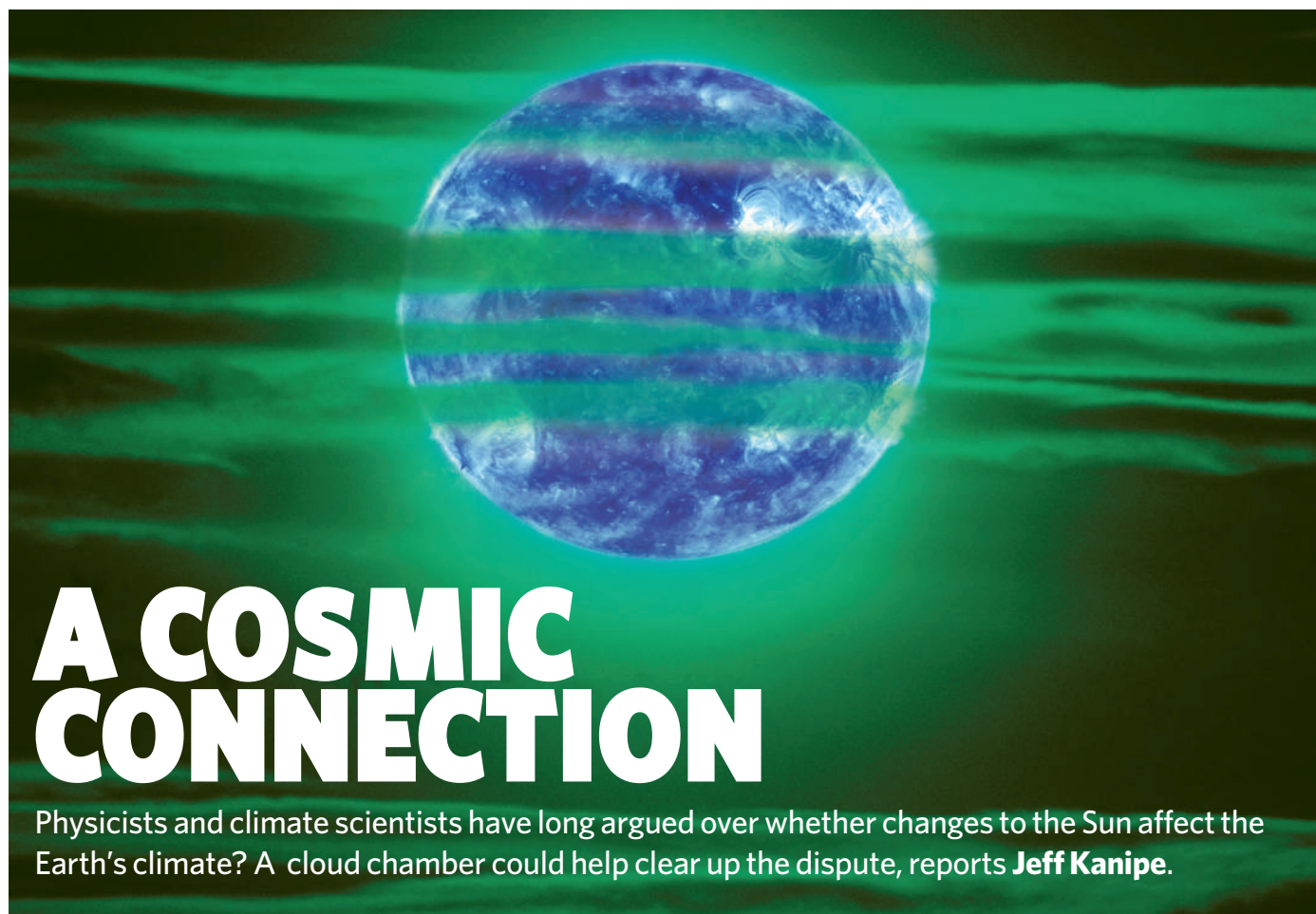
But the dodo will remain the star of the show even without DNA. Some 350 years since the last dodo picked its way through the Mauritius undergrowth, opinions on the bird continue to change. “It may be dead, but it is very much alive,” says Rijsdijk.

The dodo attracts money even now. Rijsdijk has created the nonprofit Dodo Research Foundation, hoping to fund an expedition to the Mare aux Songes next year. He plans to excavate with painstaking care, avoiding the destructive mechanical techniques used this summer. And in the years to come, there could be boxes full of dodo bones to study — and perhaps, finally, enough material to dispel the many myths that have latched onto this bird.

Henry Nicholls is a freelance writer in London.

1. Roberts, D. L. & Solow, A. R. *Nature* **426**, 245 (2003).
2. Hume, J. P., Martill, D. M., & Dewdney, C. *Nature* **429**, doi:10.1038/nature02688 (2004).
3. Cheke, A. S. in *Studies of Mascarene Island Birds* (ed. Diamond, A. W.) 5–89 (Cambridge Univ. Press, Cambridge, 1987).
4. Cheke, A. S. *Ibis* **148**, 155–158 (2006).
5. Janoo, A. *Ann. Paleontol.* **91**, 167–180 (2005).
6. Shapiro, B. et al. *Science* **295**, 1683 (2002).
7. Hume, J. P. *Hist. Biol.* **18**, 65–89 (2006).

“We have a great opportunity to reconstruct the dodo’s world.”
— Kenneth Rijsdijk



A COSMIC CONNECTION

Physicists and climate scientists have long argued over whether changes to the Sun affect the Earth's climate? A cloud chamber could help clear up the dispute, reports **Jeff Kanipe**.

The cloud chamber, invented by C. T. R. Wilson at the start of the twentieth century, opened a new microcosmos to human examination. Droplets of water coalescing around ions in a chamber saturated with water vapour provided physicists with a way of visualizing particles; tracks left by the particles were preserved like contrails across the sky (see 'The first cloud maker', overleaf). In the second half of the century, though, the technology was eclipsed — first by bubble chambers and then by a range of other devices.

Which is why it is surprising that an international team of scientists at the biggest particle-physics laboratory in the world, CERN in Geneva, Switzerland, is hard at work on a cloud chamber remarkably like those used by the physicists who founded the lab in 1954. The difference is that scientists will not use the clouds in this chamber to look for subatomic particles; they will use the particles in the chamber to try and understand clouds.

The experiment, called CLOUD (for Cosmics Leaving Outdoor Droplets), is designed to shed light on a sometimes acrimonious debate between a small number of physicists and astronomers, who believe that cosmic rays have a substantial influence on Earth's climate, and many in the mainstream climate community who don't. In CLOUD, a beam of particles from CERN's Proton Synchrotron will stand in

for the cosmic rays. And a team of atmospheric physicists, chemists and space scientists from nine countries will try to see how they affect cloud formation¹.

The CLOUD chamber has various bells and whistles that its particle-focused forerunners lacked, such as control systems to simulate parcels of air rising through the atmosphere and a 'field cage' to generate an electric field similar to that found in fair-weather clouds. The chamber can generate a range of water-vapour supersaturations (relative humidity values above 100%) that will allow droplets to form on particles between 0.1 nanometres (the size of small ions) and 100 nanometres (the size of the nuclei around which drops form in natural clouds).

Sun trap

The CLOUD prototype currently under construction will see its first subatomic particles from the synchrotron in October, according to team leader Jasper Kirkby. The machine will be fully operational in 2010, unless there are hitches with its €9.1-million (US\$11.6-million) budget. "Although this is a much simplified version of the final CLOUD experiment," says Kirkby, "we feel confident that, as this will be the first experiment of this kind on a particle beam, there will be very interesting new physics results."

One of the earliest connections between

clouds and cosmic rays was made in 1997 in the *Journal of Atmospheric and Solar-Terrestrial Physics* by Henrik Svensmark and his colleague Eigil Friis-Christensen². The two researchers were then with the Danish Meteorological Institute in Copenhagen, Denmark, and are now at the Danish Space Research Institute, also in Copenhagen. Svensmark and Friis-Christensen noted that from 1987 to 1990, global cloudiness fell by approximately 3% and that the number of cosmic rays reaching the Earth dipped by 3.5%, a fluctuation that coincided with a peak in sunspot numbers. Looking at sea and land temperatures and cosmic-ray data from 1970 to 1989, they concluded that variations in cosmic-ray intensity that depended on the sunspot cycle might be influencing global cloud cover².

The idea that changes in the Sun's activity can influence Earth's climate is a long-standing one. In the 1970s, US astronomer Jack Eddy found a link between the Maunder minimum — a dearth of sunspots in the seventeenth and eighteenth centuries, which reflects a change in solar activity — and the 'little ice age', as indicated by the width of tree rings. But the problem with such links was that the changes in the Sun's overall energy output seemed much too small to account for the change in the climate³.

The cosmic-ray hypothesis provided a way around that problem. It was not, the Danes

argued, a change in the Sun's output of photons that so affected Earth, it was a change in the way the stream of particles ejected from the Sun, the solar wind, interacted with Earth's magnetic field. The Sun's activity waxes and wanes in regular cycles, as revealed by changes in sunspots. Decades of data from balloon-borne radiosondes and particle detectors show that the number of cosmic rays reaching the Earth goes down when this activity is at its height, because the strengthened magnetic field associated with the solar wind deflects them. During quiescent periods, the solar wind weakens, allowing more cosmic rays to penetrate to Earth's lower atmosphere.

There the rays ionize atoms. They also occasionally hit the nuclei of atoms hard enough to create new isotopes such as carbon-14 and beryllium-10. These isotopes, stored in layered receptacles such as tree rings and ocean sediments, provide a record of solar activity going much further back than human observations — a record that, at some points, does seem to fluctuate in step with the climate⁴.

Clouding the issue

Svensmark's idea was that trails of ionization in the atmosphere caused by cosmic rays created condensation nuclei on which cloud droplets could form. More cosmic rays would mean more cloud condensation nuclei and either more numerous clouds, whiter clouds, or both. These clouds would increase the rate at which sunlight was reflected away from the planet and so cool it down. Svensmark and his colleagues cite studies⁵ showing that, since 1964, galactic cosmic-ray intensity has declined by about 3.7%, and that the trend indicates that this decline could have been going on since the early twentieth century. This, they argue, could account for a significant amount of the global warming that's occurred over the past 100 years⁶.

This cosmic-ray connection drew a lot of media attention for several years, but never found favour with the mainstream of climate science, which holds that the twentieth century's global warming was caused by people, not particles. To many in the community, the attention paid to Svensmark and Friis-Christensen seemed to be at best a diversion, at worst a counter-attack. The connection with the Sun was played on by organizations with connections to oil companies, such as the right-wing George C. Marshall Institute in Washington DC.

There were also questions about Svensmark's use of data. In a 2004 article published in *Eos*, Paul Damon of the University of Arizona in Tucson and Peter Laut of the Technical University of Denmark discussed several examples of what they called "unacceptable handling of observational data" by Svensmark and Friis-Christensen which exaggerated the correlation⁷.

Among several flaws, including arithmetical



errors, they noted that the cloud data that had been used originally did not represent total global cloud cover, and that when the correct data were used the correlation broke down. Svensmark began to use a different measure of cloudiness, justifying this by arguing that the new measure made more sense than the original one as something that the cosmic rays might be influencing.

CLOUD, which Kirkby has been trying to get off the ground for years, is an attempt to move beyond such arguments. It is not a completely neutral attempt — Kirkby gives talks that put a strong emphasis on the cosmic-ray interpretation of climate history, and Svensmark is a member of the CERN team.

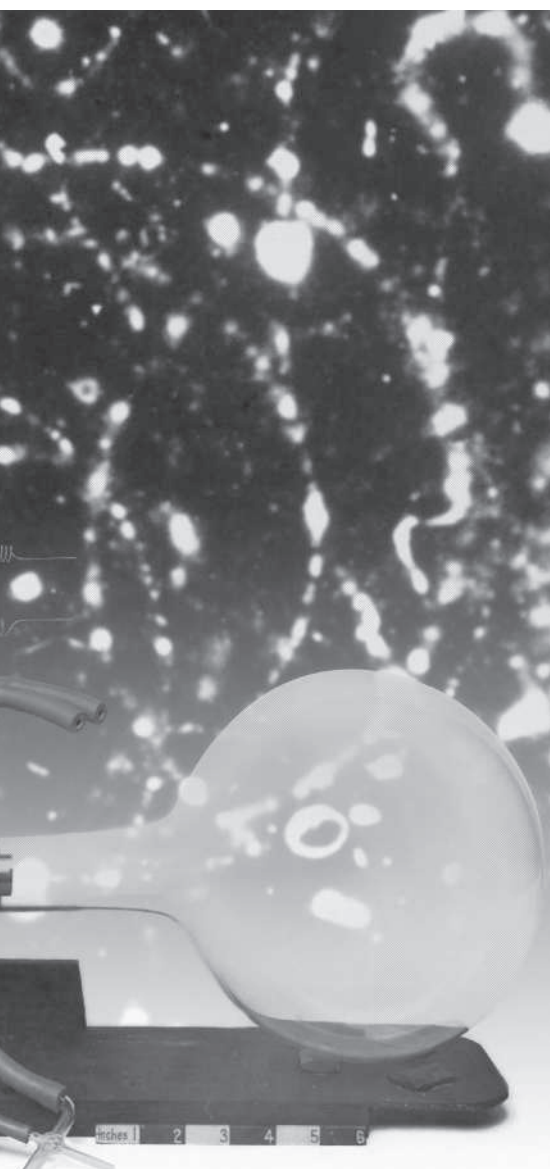
But it will produce fresh data. "CLOUD can compare processes when the 'cosmic-ray' beam is present and when it is not," says Kirkby, "and all experimental parameters can be controlled and measured."

One of those who will be watching is Nir

Shaviv, an astrophysicist at the Racah Institute of Physics in Jerusalem, Israel. Shaviv bases a belief that global warming may not be solely the result of anthropogenic causes on studies of climate over extremely long timescales — hundreds of millions of years. From an analysis he has made of 80 iron meteorites, Shaviv claims to have discovered a periodicity in cosmic-ray exposure, with a peak every 143 million years, give or take 10 million years. He says this corresponds to the mean time between four crossings of the Sun through the Galaxy's spiral arms⁸.

The spiral arms of a galaxy hold the greatest concentrations of massive stars, stars which over the course of tens of millions of years explode as supernovae. As cosmic rays are thought to be generated in supernova shockwaves, the Sun's passage through a spiral arm would mean greater cosmic-ray exposure, more clouds, and colder temperatures for Earth. Shaviv thus argued that these periodic crossings and the high cosmic-ray flux they bring with them correspond to four major ice-age epochs on Earth. If this could be confirmed it

"We feel confident that there will be very interesting new physics results."
— Jasper Kirkby



THE FIRST CLOUD MAKER

In September 1894, Charles Thomas Rees Wilson, an expatriate Scot, climbed to the summit of Ben Nevis, the highest mountain in his homeland. He was fascinated with the beauty of coronae and the dazzling, rainbow-like rings encircling human-shaped shadows called Brocken spectres. He returned to the Cavendish laboratory in Cambridge keen to create controlled clouds and so study such effects.

Wilson designed a glass tank with a piston tightly inside it (see left). He misted the air in the tank with water until it was saturated, then rapidly pulled out the piston. This expanded and chilled the air, causing the water vapour to collect around dust particles and form a small cloud. As long as the piston was pulled out far enough, clouds would form even if the air inside the chamber was thoroughly filtered. Wilson suspected, but couldn't prove, that the water droplets were forming around electrically charged ions.

In February 1896, Wilson projected a beam of X-rays into his cloud chamber, and the mini clouds formed even faster, thus lending support to his theory. A year later, after Joseph John Thomson discovered the electron, it became clear that the X-rays ionized the air molecules and the electrons and ions created cloud nuclei. Wilson continued his experiments for another decade, during which he discovered that particles colliding with atoms in the chamber left wispy tracks of ions in their wake. When subjected to a magnetic field, the curve of a track indicated whether the ion was negatively or positively charged. In 1927, Wilson was awarded the Nobel prize in physics for his cloud chamber. By then, though, he had left particle physics to study the atmosphere. When CERN was founded in 1954 he was retired but still active, working on a theory of thunderclouds. **J.K.**

crossings to ice-age epochs or taints the idea that cosmic rays may influence Earth's climate. "I think the findings in our paper are still valid and add weight to arguments about the possible cosmic ray-climate connection advocated by Shaviv and colleagues." Shaviv and Veizer, too, stand by their findings.

Rahmstorf, who led the charge against Shaviv and Veizer, agrees that solar variability has an influence on climate, but he is adamant that galactic cosmic rays cannot explain the global warming of the past decades. "We know that the observed climate response is the sum of several forcing factors, including the Sun, which is probably responsible for some part of the warming up to about 1940, but not for warming after that. The Sun simply shows no trend there, and neither does the cosmic-ray flux."

Perhaps the irony here is that the current emphasis on global warming is encouraging people from outside the field of climate studies to try out ideas, perhaps exaggerating their promise. At the same time those within the climate community are mindful of lobbies that seek to discredit their work, and believe enough doubt has been cast on the cosmic-ray theory to discredit it, if not ignore it completely.

Yet for a connection between cosmic rays and climate to be interesting, it does not have to account for the already well-explained climate history of the past 100 years. Even a small effect, to which Earth is only sensitive under some conditions, would be an exciting find. The CLOUD experiment does not have to overturn the consensus of the world's climatologists to be a success; it just has to throw a little light on some physics. "In a nutshell," says Kirkby, "we want to go after the microphysics between a cosmic ray and a cloud droplet or ice particle. How significant are they in the atmosphere, or in certain parts of the atmosphere?"

And even if the results prove undramatic, or inconclusive, they will still carry a certain sense of historical redress. Wilson did not set out to invent a tracker for particles; he wanted to understand the nature of clouds. For his technology to be returned to that aim more than a century on would undoubtedly have pleased him.

Jeff Kanipe is a science writer based in Maryland.

1. The Cloud Collaboration; preprint at <http://arxiv.org/abs/physics/0104048> (2000).
2. Svensmark, H. & Friis-Christensen, E. *J. Atmos. Solar-Terrest. Phys.* **59**, 1225-1232 (1997).
3. Foukal, P., Fröhlich, C., Spruit, H. & Wigley, T. M. L. *Nature* **443**, 161-166 (2006).
4. Neff, U. *et al. Nature* **411**, 290-293 (2001).
5. Lockwood, M., Stamper, R. & Wild, M. N. *Nature* **399**, 437-439 (1999).
6. Marsh, N. & Svensmark, H. *Space Sci. Rev.* **94**, 215-230 (2000).
7. Damon, P. E. & Laut, P. *Eos* **85**, 370-374 (2004).
8. Shaviv, N. J. *Phys. Rev. Lett.* **89**, 051102 (2002).
9. Shapley, H. *J. Geol.* **29**, 502-504 (1921).
10. Gies, D. R. & Helsel, J. W. *Astrophys. J.* **626**, 844-848 (2005).
11. Shaviv, N. J. & Veizer, J. *GSA Today* **13**, 4-10 (2003).
12. Rahmstorf, S. *et al. Eos* **85**, 38-41 (2004).
13. Jahnke, K. preprint at <http://arxiv.org/abs/astro-ph/0504155> (2005).

would link changes in the solar neighbourhood directly to climate. This idea was first considered by the great American astronomer Harlow Shapley, who speculated in 1921 that an 80% change in solar radiation due to the irregularity of interstellar clouds "would completely desiccate or congeal the surface of the Earth"⁹.

Shaviv's findings have found favour with some astronomers, such as Douglas Gies, an astrophysicist at the Center for High Angular Resolution Astronomy at Georgia State University in Atlanta.

Based on the Sun's current position as determined from the distances and motions of nearby stars, Gies and a colleague, John Helsel, tracked the path of the Solar System through the Galaxy for the past 500 million years. Their findings indicate that the Sun could indeed have crossed four times through segments of the spiral arms, each passage corresponding with four major ice ages¹⁰.

Again, climate scientists disagree. They think that the history of Earth's glaciations is pretty well accounted for by changes in the amount of carbon dioxide in the atmosphere. And in this

instance, too, the climatologists contend that the astrophysicists' claims, specifically Shaviv's, are unfounded and even misleading. A paper¹¹ coauthored by Shaviv and Ján Veizer, now an emeritus professor at the University of Ottawa in Canada, was dressed down by no less than 11 scientists led by Stefan Rahmstorf of the Potsdam Institute for Climate Impact Research in Germany. The scientists contended among other things that Shaviv and Veizer had misinterpreted the cosmic-ray data derived from their meteorite samples¹².

A century on

A follow-up analysis of the same data set a year later by Knud Jahnke, from the Max-Planck-Institute for Astronomy in Heidelberg, Germany, showed no evidence of significant changes in cosmic-ray exposure with a 143-million-year period — or any other between 100 million and 250 million years¹³. Again, there was a heated to and fro (see Shaviv's blog at www.sciencebits.com).

Gies doesn't think the controversy changes Shaviv's findings relating Earth's spiral-arm

Investigator-initiated trials are more impartial

SIR — Your Naturejobs Special Report “Trial blazers” (*Nature* **442**, 480–481; 2006) highlights the demand for investigators adept at planning, running and managing clinical trials for the pharmaceutical industry.

The industry’s share of total biomedical research increased from nearly a third in 1980 to nearly two-thirds in 2000. Nevertheless, a report documenting combined data from more than 1,100 studies showed that industry-sponsored clinical trials are “significantly more likely to reach conclusions that were favorable to the sponsor than were non-industry studies”, possibly because of publication bias or selection of an inappropriate comparator to the drug being evaluated (J. E. Bekelman, Y. Li and C. P. Gross *J. Am. Med. Assoc.* **289**, 454–465; 2003).

Efforts being made by some academic institutions and peer-reviewed journals to control financial conflicts include the disclosure of financial interests by researchers in academia and as authors in peer-reviewed journals; statistical analyses performed by independent academic researchers; and research protocols and results from clinical trials made fully available in registries accessible to the public.

Given that the trend towards industry’s increased share of biomedical research will undoubtedly continue, further evaluation of differential bias between investigator- and industry-initiated contracts for clinical trials within academic medical centres is needed. Of particular note in this respect is the fact that some investigator-initiated trials are funded entirely by industry, whereas others are funded via peer-reviewed grants or foundations.

Within academic institutions, for example, trials funded by industry but initiated by the investigators typically have 10–20% lower overhead costs. They tend to restrict industry sponsors to ensuring adherence to good clinical-practice guidelines and to checking data for serious adverse events, notation errors and omissions. Hence industry has very limited influence on the conduct of the trial. In contrast, trials that are both initiated and funded by industry within academic institutions have higher overhead costs, usually in the range of 30–40%, paid to the academic institution. Such trials allow greater involvement by the industry funder, for example in writing the protocol, training and assisting investigators, and helping in data collection and analysis.

Academic investigators should be encouraged to lead investigator-initiated clinical trials, thereby limiting industry involvement and reducing any bias that may creep into trial protocols, conduct and

results. This should result in better patient care via increased knowledge about the safety and efficacy of the drug therapy under evaluation.

Bradley C. Johnston, Sunita Vohra

Department of Pediatrics, University of Alberta, Edmonton, Alberta T6G 2E1, Canada

Biochar trials aim to enrich soil for smallholders

SIR — Your recent News Feature “Black is the new green” (*Nature* **442**, 624–626; 2006) accurately summarized the origin, discussions and goals of the Terra Preta Nova group at the recent World Congress of Soil Science in Philadelphia.

However, one might be left with the impression that the biochar initiative is solely directed towards agribusiness applications. From the start, this has certainly not been the case. Indeed, innovative biochar field trials involving a variety of crops are currently being conducted in Amazonia by researchers in Embrapa, the Brazilian agricultural research corporation, and INPA, Brazil’s national institute of Amazonian research. These trials are specifically designed for implementation by smallholders, who comprise most of the world’s farmers. At the same time, the Brazilian researchers are seeking answers to the questions raised in Philadelphia about the native *terra preta* of Amazonia, even as US and European researchers are taking the concept in new directions.

William I. Woods*, Newton P. S. Falcão†, Wenceslau G. Teixeira‡

*Department of Geography, University of Kansas, Lawrence, Kansas 66045, USA

†INPA/CPCA/Solos e Nutrição de Plantas, Av. André Araújo, Noo 3936, Bairro Petrópolis, CEP. 69011-970, Caixa Postal 478, Manaus - AM, Brazil

‡Embrapa Amazônia Ocidental, Rod. AM 010 - Km 29, CEP. 69011-970, Manaus - AM, Brazil

Biochar and biofuels for a brighter future

SIR — Your excellent News Feature “Black is the new green” (*Nature* **442**, 624–626; 2006) refers to President Bush’s announcement that \$150 million is to be put into converting cellulosic precursors to ethanol — a very modest amount, considering the plight of petrochemicals. I would like to mention here some progress already being made in producing biofuels and platform chemicals from cellulose and hemicellulose precursors.

Almost quantitative yields of levulinic acid — an excellent platform chemical that gives

rise to fuel additives, polymers and plastics, and numerous essential chemicals — are obtained from these precursors in the Biofine process (the work of Stephen Fitzpatrick of Biofine Renewables LLC in Waltham, Massachusetts). A 300-tonne-a-day Biofine plant is nearing final commissioning at Caserta, Italy. Similarly, commercial yields of fermentable levoglucosan from cellulose, and furfural from the pentoses in hemicelluloses, have been obtained in pilot studies using the Convertech process, led by Ken Scott of Scott Convertech in Christchurch, New Zealand.

The biomass to fuel such processes will revolutionize agriculture. It will also give rise to sustainable chemical industries, as oil reserves become depleted and oil prices are pushed up by demand, depletion and political dictates. The biorefinery waste is a char. This has a calorific value similar to that of bituminous coal, and our studies indicate that it has, like the Terra Preta de Índio chars, the potential to be an excellent soil ameliorant. And looking at yet another innovation: the future will be bright if we base it on Carboleum, or ‘oil from carbohydrate’ — the brainchild of Austin Darragh at the University of Limerick — as we say farewell to petroleum.

M. H. B. Hayes

Chemical and Environmental Sciences, University of Limerick, Limerick, Ireland

Science networks spread throughout the Arab world

SIR — I agree with your Editorial “Science in the Arab world” (*Nature* **441**, 1027; 2006) that the move towards better funding of research and scientific activities in the Arab world is welcome, even though this move remains disproportionate to the wealth of the region.

Although the Arab Science and Technology Foundation is doing good work, it is important not to convey the picture that it is the only organization dealing with research in the area. There is great interest in research among governmental and non-governmental organizations, in addition to individual efforts. The eastern Mediterranean region of the World Health Organization is very active in this respect. In addition, there is a surge in the number of research networks and online journals, particularly in the areas of medicine and clinical research.

A. Abyad

Health Policy and Regulation, General Authority for Health Services for the Emirate of Abu Dhabi, PO Box 5674, Abu Dhabi, United Arab Emirates

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter.

BOOKS & ARTS

Our place in the Multiverse

Was the appearance of a Universe that can support life inevitable?

Many Worlds In One

by Alex Vilenkin

Hill & Wang: 2006. 248 pp. \$24

Joseph Silk

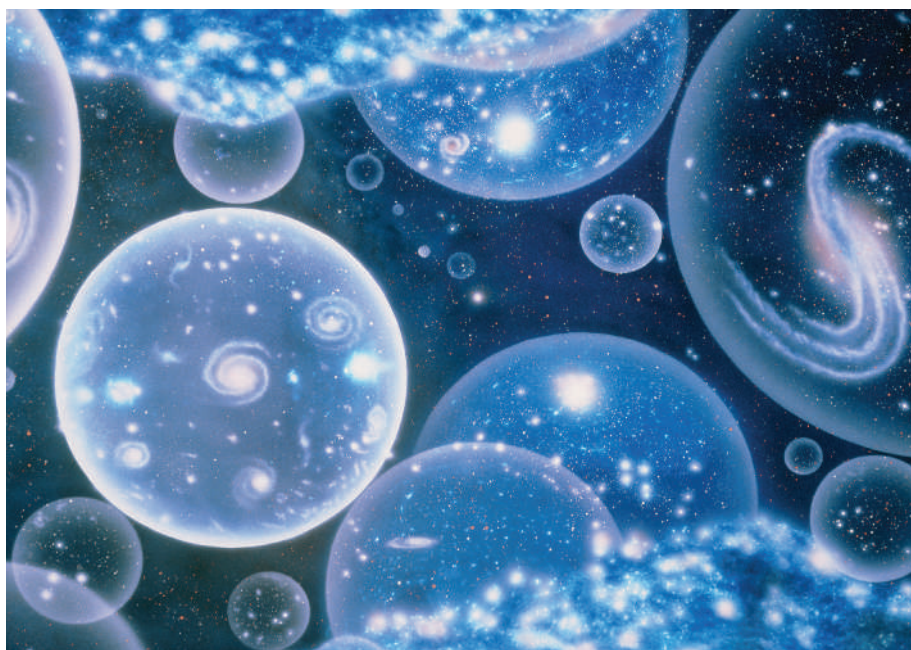
The smallest person in the world, an Indian called Gul Mohammed, had a height of just 57 centimetres. The tallest, an American called Robert Wadlow, measured 2.72 metres. But the observed distribution of human heights fills only a small part of the range in between. Why? The average is about 1.63 metres, a respectable height for much of the world's population. Where are the giants of fairy-tale fame? And where are the Lilliputians?

We know the answers. Our genes control the supply of the growth hormones that spur our bones to elongate. Genetic abnormalities can reduce the abundance of these hormones in dwarfs, and oversupply leads to gigantism. The environment, most notably nutrition, also plays a part, and gravity constrains our height — we are taller after sleeping, and astronauts gain height in space. Genetic evolution, with help from physical constraints, has narrowed the height range to the observed distribution.

Here we can reconcile observation with theory using known physics and biology; there is no need to invoke another explanation, such as a Grand Designer. But does the same kind of reasoning apply when scientists discuss our place in the Universe? In his stimulating new book, *Many Worlds in One*, cosmologist Alex Vilenkin invokes the anthropic principle in his interpretation of the Universe we observe: it is the way it is because we are here to observe it.

There are many possible universes that are inhospitable to our existence. The latest theories of quantum gravity count some 10^{500} realizations of the universe, in which the various fundamental constants of nature differ. In this Multiverse, all universes are equally real, although we can only hope to explore our own one. Given the staggering array of alternatives, it is exceedingly unlikely that our observed universe should even exist.

Take the mystery of dark energy, for instance, which dominates the energy density in the Universe. Our best theories predict a value for the amount of dark energy that is too large by a factor of 10^{120} . It is a tautology to assert that our existence selects an appropriate universe from the ensemble of all universes. After all, we can only observe a universe of a certain size,



Bubble vision: theories of quantum gravity suggest that the Multiverse may contain 10^{500} universes.

old enough for stars and planets, and for life to have developed. But it is physics, or at least metaphysics, to state, as the physicist Robert Dicke first did, that the Universe must be old enough for stars to have synthesized carbon, a necessary condition for our presence. It is one further logical step to assert that the values of all of the fundamental constants of nature, which may vary throughout the Multiverse, are determined by our presence. This is the anthropic principle in its weakest form. It is simply observational selection, with the caveat that our presence is not guaranteed.

A strong version of the anthropic principle claims that intelligent life is inevitable somewhere in the Multiverse. But let us put that aside, if only because strong anthropic arguments are weakened by the inclusion of a possibly infinite age for the Universe. A great deal can happen over a long time in a universe that perpetually renews itself by eternal inflation. The weak anthropic principle, preferred by many of my colleagues, selects only the small subset of 'pocket universes' within the Multiverse that allow galaxies to form and life to develop. There is then a high probability of finding only a small but non-zero value for dark energy today, which is what we observe.

At least three rival hypotheses could explain

the values of the fundamental constants of nature. First, the selection could have been made by a Grand Designer. This has great appeal to proponents of the intelligent design of the Universe. Vilenkin argues forcefully that there is no need to invoke such a concept, although ultimately it reduces to a question of personal belief. The second option appeals to currently unknown physics. The height distribution of human beings can be understood by known rules, so there is no need to invoke another explanation; perhaps we simply do not yet know the rules for navigating in the Multiverse. It may be that it takes an infinite time to populate the plenitude of potential universes. The ultimate voyage through the quantum foam that characterizes the Multiverse to arrive in our Universe may take so long that it could only have happened once. If so, it makes no sense to talk of probabilities for our Universe to appear.

The third option, and to my mind the most likely, is that there was no selection at all: we are here because we are here. This is what must happen in an infinite Multiverse. Some versions of quantum gravity appeal to the complexity of the initial conditions to assert that there were an infinity of landscapes and universes in the Multiverse. If this were the case, the game is

J. BAUM/SPL

over. The dice were rolled and our Universe was inevitable, somewhere in the Multiverse. And here we are.

Remarkably, we can test this hypothesis. Future experiments will measure the curvature of space with exquisite precision. If the curvature turns out to deviate from flatness, we would come to a conclusion unprecedented in human thought. A slightly closed Universe would prove the finiteness of space. A slightly open Universe would go a long way towards demonstrating that space is infinite, at least in

standard cosmologies. If this were the case, we would no longer need to invoke any version of the anthropic principle; it would simply be redundant.

I thoroughly recommend *Many Worlds in One*. Vilenkin has made some major contributions to the Multiverse hypothesis. Here he illuminates the current issues with clarity and elegance, yet the stories he tells are accessible to non-specialists. ■

Joseph Silk is in the Department of Physics, University of Oxford, Oxford OX1 3RH, UK.

Back on track

Out of the Woods: Tales of Resilient Teens

by Stuart T. Hauser, Joseph P. Allen & Eve Golden

Harvard University Press: 2006. 327 pp. \$27.95

Ann S. Masten

Stories of individuals overcoming great obstacles to succeed in life no doubt cast their spell on audiences long before they were recorded in books. The power such tales hold will captivate readers of *Out of the Woods*, an accessible book on resilience by two leading clinical scientists in developmental psychiatry and psychology, Stuart Hauser and Joseph Allen, and independent scholar Eve Golden. Drawing on data from a unique study of 67 teenagers hospitalized for serious behavioural and emotional problems, the authors examine how nine of the troubled teens managed to get 'out of the woods' and on to positive developmental pathways as adults. Their voices emerge from striking interview narratives that span a dozen years, from the time of initial hospitalization (when the future looked bleak indeed) into early adulthood. The achievements and gaps of the resulting book reflect the progress as well as the difficulties of 'resilience science'.

It is intriguing to see a return to case narratives at this particular time in the history of resilience research. Compelling cases often inspired the pioneers who launched the systematic study of human resilience in the 1960s and 70s. Much was learned over the ensuing decades, despite the new field suffering growing pains. A range of diverse studies yielded surprisingly congruent findings about potential protective factors, or 'what makes a difference', although frustratingly few details emerged about how these factors actually work — a situation concisely summarized in the opening chapters.

The heart of the book is focused intensively on the lives of four resilient young people, who are compared and contrasted with each other and with their less adaptive peers. Anyone concerned about young people will find their stories thought provoking. The book reveals the changing voices and perspectives of each

individual and shows, in the words of the authors, "how resilience can kick in before it shows". Some of the interview narratives also provide an entertaining and illuminating portrait of unconventional young people thwarting the conventional strategies of their interviewers as the teens struggle to hold their own against family adversity, the indignities of a locked inpatient unit, and the stigma of treatment for mental illness. There are fewer words from the more typical patients; readers must rely on commentary by the authors about the differences.

The case studies emphasize three fundamental protective factors, each a long-standing member of the resilience hall of fame. The first is the presence of positive relationships, encompassing the attachment system and the human capacity for recruiting and forming lasting bonds with parent figures, partners or mentors. The second is agency, the capacity and confidence people have for steering their own lives, and the reward systems that motivate efforts to overcome obstacles and control what is happening in one's life. The third protective factor is reflection on one's self and situation, and the requisite cognitive systems for directing attention and thought. Readers will find other well-established protective factors in these case reports, including optimism about the future and the belief that life has meaning, access to positive institutions, and the advantages of

certain cultural or socioeconomic positions.

Largely missing from these vivid accounts, as well as from the first generation of resilience science, is the role of context and multiple levels of analysis. These cases offer a rare and fascinating glimpse into the lives of hospitalized adolescents who have recovered, yet it is difficult to apply the lessons from their lives to other teens and situations. As the authors point out, many details about the developmental histories of these individuals remain unknown. The general role of developmental changes, in person and situation and the way these interact, is not emphasized. Researchers now speculate that the transition to adulthood provides a window of opportunity for late bloomers, when motivation, cognition and opportunity converge to make positive change possible. Brain development enhances the capacity for evaluating, planning and redirecting one's own life. At the same time, young people encounter new and transformational situations, in the form of higher education, job training, military service, religious service or courtship. There is a new wave in resilience science, based on efforts to understand processes across disciplines, levels of analysis and species (humans and animals). Scientists are beginning to integrate recent advances in genomics, neuroscience and the modelling of dynamic systems into the body of knowledge on resilience.

The book serves as a powerful reminder of the phenomenon of resilience and the compelling rationale for understanding resilience well enough to facilitate it. All these young people have suffered, and only 13% (so far) have made it out of the woods. Their resilience inspires hope, just as individual cases inspired pioneers 40 years ago to hope that science could improve the odds for resilience or prevent children from getting into the woods in the first place. As resilience science matures, the prospects for realizing these goals are improving. Meanwhile, it is important to remember that the woods continue to fill with young people who have lost their way. ■

Ann Masten is distinguished McKnight university professor of child development, 51 East River Road, Minneapolis, Minnesota 55455-0345, USA.



Down but not out: research is progressing on how to help young people overcome their problems.

N. GOELBAS/GETTY

A mathematical tonic

Dr Euler's Fabulous Formula: Cures Many Mathematical Ills

by Paul Nahin

Princeton University Press: 2006. 404 pp.
\$29.95, £18.95

Timothy Gowers

One of the effects of the remarkable success of Stephen Hawking's *A Brief History Of Time* (Bantam, 1988) was a burgeoning of the market for popular science writing. It did not take long for mathematics to get in on the act, especially when Andrew Wiles captured the public imagination by proving Fermat's last theorem. Now scarcely a month goes by without a new popular mathematics book. Some are about famous unsolved problems, others are about notable numbers such as zero, π or e , and still others, including *Dr Euler's Fabulous Formula* by Paul Nahin, are about important equations.

Mathematics is not easy to explain, and many authors writing popular mathematics go out of their way not to lose their readers, perhaps bearing in mind Hawking's famous dictum that each equation included in a book will halve the sales. But what if, like most readers of *Nature*, you already know some mathematics? Then a book that avoids mathematical symbolism is likely to be frustratingly vague. Fortunately, however, among the books that have appeared recently, there are some that might be called 'semi-popular': they are still rather informal but are aimed at a more mathematically sophisticated audience. *Dr Euler's Fabulous Formula* is a welcome addition to this category.

The hero of the book is Leonhard Euler (1707–83), who was one of the most prolific mathematicians ever, continuing to publish large numbers of papers even after going blind in 1766. The "fabulous formula" of the title is $e^{i\pi} = -1$, together with the more general $e^{i\theta} = \cos\theta + i\sin\theta$. These equations astonish each new generation of mathematicians that encounters them: without generous helpings of hindsight, who would ever have suspected there was such a succinct equation involving e , i and π ? Nahin is not particularly concerned with the philosophical problems associated with complex numbers, which are discussed in his earlier, less technical book *An Imaginary Tale: the Story of $\sqrt{-1}$* (Princeton University Press, 1998). Rather, he takes Euler's formula as his starting point and demonstrates the wide range of uses to which it has been put.

As an engineer, Nahin's perspective on the formula is different from that of a typical

mathematician. This is mostly a good thing: not many mathematicians know much about signal processing or the importance of complex numbers to engineers, and it is fascinating to find out about these subjects. Furthermore, it is likely that a mathematician would have got hung up on details that Nahin cheerfully, and rightly, ignores. However, his background has the occasional disadvantage as well, such as when he discusses the matrix formulation of de Moivre's theorem, the fundamental result



Leonhard Euler linked some fundamental mathematical constants.

that $(\cos\theta + i\sin\theta)^n = \cos(n\theta) + i\sin(n\theta)$. This is a one-line consequence of the addition formulae for sines and cosines, but his proof takes up an incredible six pages and uses, quite unnecessarily, the Cayley–Hamilton theorem, which he proves in two dimensions by brute force. Even then the proof is not complete: at one point he says he will "refer you to any good book on linear algebra for a formal proof". The less experienced reader will emerge from these pages thinking that de Moivre's theorem is a dauntingly difficult result, which is a pity, as it isn't.

Sometimes Nahin reveals his engineer's instincts by dispensing with formal proof and using numerical simulations instead. In places this works well — if a result is plausible a proof isn't always needed in a book like this — but sometimes it can leave a mathematical

reader nonplussed. For instance, he discusses a problem in which a mouse runs at constant speed round a circle, and a cat, starting at the centre of the circle, runs at constant speed directly towards the mouse at all times. Does the cat catch the mouse? Yes, if, and only if, it runs faster than the mouse. In the middle of the discussion Nahin suddenly abandons the mathematics and puts a discrete approximation into his computer to generate some diagrams. His justification is that an analytic formula for the cat's position is hard to obtain, but actually the problem can be neatly solved without such a formula (although it is good to have his diagrams as well).

A question Nahin does not engage with is whether complex numbers are indispensable to his arguments. Such is his enthusiasm for complex numbers that he writes as though they always are. But there is a big difference between an argument that uses both addition and multiplication of complex numbers and one that uses just addition (in which case the complex numbers are just a convenient notation for the real plane). For example, he uses them in his analysis of the cat-and-mouse problem above, but he needn't have done. Elsewhere he explicitly says that he has used complex numbers in a fundamental way, when he hasn't. For instance, to demonstrate that a certain quadratic equation has no real solution, he points out that the familiar formula for the roots shows that they are both complex. This is not a true use of complex numbers, because to derive the formula for the roots one must complete the square, and then the true reason for the lack of real roots emerges: the square of a real number is never negative.

On the whole these defects are not serious, however. Indeed, mathematicians may well find them thought-provoking and others will not notice them. On the positive side, Nahin includes gems from all over mathematics, ranging from engineering applications to beautiful pure-mathematical identities. Most of his topics lie just beyond the periphery of a typical mathematics course: they are facts, such as the irrationality of π , that you may have heard of but never had explained in detail. It would be good to have more books like this. ■

Timothy Gowers is at the Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK.

Correction

In the review of the book *Soils and Societies: Perspectives from Environmental Histories* (*Nature* **442**, 985; 2006), the name of one of the editors was misspelled. The editors were J. R. McNeill and Verena Winiwarter.

AKG IMAGES

A global vision

The Ecological Theatre and the Evolutionary Play

by George Evelyn Hutchinson
Yale University Press, 1965

Alison Jolly

"For the purposes of this seminar, let us define ecology as the study of the Universe." Evelyn Hutchinson's hooded blue eyes surveyed the graduate students assembled in Yale University's zoology department. We were products of the complacent 1950s and felt we were embarking on a voyage far from known shores, piloted by an ultra-eccentric, ultra-English don.

Hutchinson told us about Arrhenius's notion that industrial carbon dioxide might lead to global warming. He mooted the idea that fertilizer affects the world's phosphate cycles, and wondered if it mattered that when the world's cows break wind, they emit methane. Apparently we lived in a 'biosphere', a concept derived from Vladimir Vernadsky's publications in French and Russian. Hutchinson zoomed from global biogeochemistry to animalcules so small that brownian motion rules their lives. He suggested that each living species occupies a multidimensional niche, perpetually jostled by the impact of other species, but mathematically tractable if you treat the niches' dimensions as fine-grained scales of space and time. The web of ideas gleamed like spider-threads stretched between stars.

In 1965 Hutchinson published a slim book called *The Ecological Theatre and the Evolutionary Play*. He began by speculating whether life was probable on an Earth-like planet. He described the biosphere, the multidimensional niche, then lakes and ponds as microcosms. To help us understand these microcosms, Hutchinson embraced a range of knowledge from mud chemistry to thermal stratification, the relative growth rates of water-beetles, and the history of the surrounding agriculture. The details of this work he later covered in the four massive volumes of his *Treatise of Limnology* (Wiley, 1975–1993).

The Ecological Theatre and the Evolutionary Play instead danced onwards to the evolution of human intelligence, the phallic symbolism of Neapolitan coral baby-charms, and the importance of randomness in human evolution and art. The book ended with an apparently trivial essay entitled "The Cream in the Gooseberry Fool". (It didn't stop to explain that a fool

is an English recipe of fruit blended with dollops of heavy cream.) Here, Hutchinson dwelt on the Reverend Gilbert Raynor's passion for wing variation in the gooseberry moth, *Abraxis grossulariata*. Raynor's moth-breeding records finally became one small foundation of biochemical genetics. An onlooker described the 1907 sale of Raynor's collection as "gooseberry fool, mellowed...by some of that cream which regards these fine aberrations as matters of scientific interest and not, as so many say, postage stamps".



George Evelyn Hutchinson foresaw many of the problems that are now known to affect the biosphere.

Hutchinson's ecological theories were widely acclaimed. His clutch of honorary doctorates culminated in the scarlet robes and Latin oration of the University of Cambridge, where his father had been master of Peterhouse College. But Hutchinson's colleagues at Yale were often exasperated, especially when he promoted heresies, such as Lynn Margulis's wild claim that eukaryote cells derive from symbiosis.

His students were chronically bemused. In our examinations he set us impossible questions: "Imagine an utterly isolated island in a constant environment. How big would it have to be to permit indefinitely ongoing evolution?" The students ricocheted off in all directions: towards limnology, or calculating species-area ratios on islands, or landscaping bits of the Amazon jungle to assess the minimum critical size of forest ecosystems, or analysing competition and symbiosis between rainforest trees, or studying maternal behaviour in goats and intelligence

in primates. (Hutchinson once sent a juvenile bushbaby hopping down the middle of the table in a department faculty meeting, convinced that even molecularly oriented colleagues would be enchanted by the creature. Unfortunately they weren't.)

Is science the delight in bushbabies, pond-beetles and aberrant moth-wings, or the understanding that we need to guide our species through the current geological era, the Anthropocene? Hutchinson's larger vision was that the Universe is an interacting whole. James Lovelock has taken this much further with the concept of 'Gaia', which regulates Earth's systems so life can continue. Hutchinson was more cautious: he claimed only that in the macrocosm, as in the microcosm, everything is connected.

That vision now seems the basis of human survival. We have global warming. Global plagues. Global crises of water, land use and loss of biodiversity. Global, although certainly not free, trade. (I note that neither ring-tailed lemurs nor chimpanzees believe in free trade, and find it odd that many economists seem to.) John Maynard Smith and Eörs Szathmáry in their book *The Major Transitions in Evolution* (Freeman, 1995) explained the principles underlying the evolution of each of four great steps of biological synthesis: proto-life, eukaryotic cells, multicellular bodies,

and the societies of mammals and insects. Arguably we are now approaching a fifth major transition of evolutionary integration: is humanity on the way to a species-wide society that recognizes its dependence on a species-wide environment? That couldn't evolve by darwinian natural selection, but it could develop by the purposeful use of human (and cyber) intelligence.

Our bounded life span and our eyesight that stops at the garden wall are in direct contrast with the need for a vision of the biosphere that extends over continents and centuries — and in the other direction, down to the single-celled rulers of biogeochemistry. One of Hutchinson's impossible examination questions was: "Discuss the effect of man's size on his view of the Universe." At the time, no student dared tackle an answer. All of us, though, still wish we had.

Alison Jolly is in the Department of Biology and Environmental Sciences, University of Sussex, Brighton BN1 9QG, UK.

YALE PEABODY MUSEUM

NEWS & VIEWS



EVOLUTIONARY BIOLOGY

Human brain gene wins genome race

Chris P. Ponting and Gerton Lunter

The differences in brain size and function that separate humans from other mammals must be reflected in our genomes. It seems that the non-coding 'dark matter' of genomes harbours most of these vital changes.

The human brain is supposed to set us apart from other animals. If so, our genome must retain the imprint of our brain's recent evolution. So which parts of our genome have seen the most change, and are these genomic innovations linked directly to our unique brain structure and function? On page 167 of this issue, Pollard *et al.*¹ describe how they have clocked the speed at which various human genome regions have changed in recent times*. The clear winner of this race is human accelerated region 1 (HAR1), part of an RNA gene whose pattern of expression is suitably poised to influence the migration of neurons in the developing cortex. The authors' second and equally important finding is that all but two of the most-accelerated regions lie outside protein-coding sequences — in the enigmatic 'dark matter' of the human genome.

"Always remember that the one true, certain, final, and all-important difference between you and an ape is that you have a hippopotamus major in your brain, and it has none." Thus wrote Charles Kingsley in *The Water Babies* (1863), lampooning Richard Owen's mistaken insistence that humans alone possess a

'hippocampus minor' in their brains.

Time and again, humans' sense of cognitive superiority over other primates has failed to find a solid foundation in structural variations of the brain. The brains of humans and chimpanzees are anatomically not so different², except in scale. About two million years ago, the hominin brain began to enlarge until, in modern times, it has become about three times larger than that of chimpanzees. Size matters, but it is equally likely that alterations in cellular structures contribute to the cognitive differences between the two species. All these changes must have left their impression on the human genome. Yet evidence directly linking DNA differences to anatomical or behavioural differences between these two species has been, at best, fragmentary³.

Geneticists have long laboured under the assumption that it is the DNA sequences that code for proteins that must have borne the brunt of adaptive change³. This notion has continued despite the scarcity of coding differences between chimpanzees and humans⁴, and the recent appreciation that a greater proportion of apparently functional DNA sequence lies outside protein-coding sequence than inside^{5,6}. What was needed, instead, was an unbiased genome-wide scan to pinpoint the few regions

— of whatever type — where the DNA has remained essentially static over tens and hundreds of million years in diverse species, yet which, in the past few million years of human evolution, have altered especially rapidly.

Such is Pollard and colleagues' computational scan of the human genome¹. Their study reveals a set of 49 regions (HAR1–HAR49), each with a sequence that is highly evolutionarily conserved among many mammals, but that has diverged rapidly in humans since our last common ancestor with chimpanzees. The fastest among them, HAR1, has accrued 18 changes in sequence in this time, when only one or no substitutions would be expected to occur by chance.

How might these extraordinary changes be linked to the human brain's increased cognitive capabilities? The first clue came from the finding that *HAR1F*, one of two RNA genes containing HAR1, is expressed in the developing neocortex in the brains of humans and in those of another primate, the crab-eating macaque. This is intriguing, as the neocortex is most often associated with higher cognitive functions. Specifically, *HAR1F* is expressed in human Cajal–Retzius cells, for which a crucial role in redirecting migrating neurons has long been suspected. HAR1 thus emerges as a strong

*This article and the paper concerned¹ were published online on 6 September 2006.

candidate for a determinant of innovative function in the human neocortex.

However, other explanations for the rapidity of HAR1 evolution should not be discounted. In particular, *HAR1F* is also expressed in the ovary and testis of adult humans, and sexual selection of genes expressed in these tissues has often driven unusual sequence changes⁷. Furthermore, the authors show that HARs are often associated with regions that undergo a high rate of recombination — the process by which an offspring obtains a blend of parental genes. Recombination, and its associated process, biased gene conversion, are thought to favour the inclusion of G and C nucleotides over the other two possible nucleotides, A and T (ref. 8). As all of the nucleotide substitutions observed in HAR1 are of this type, high (and biased) mutation rates might explain part of the rapid evolution of HAR1. Nevertheless, this process cannot explain the authors' other observations, such as the pairs of substitutions that together further stabilize the structure of HAR1 RNA. Clearly, much work on the function and evolution of HAR1, and indeed of all HARs, remains to be done.

Pollard *et al.* have introduced a powerful approach for detecting the signature of adaptation in the genomic sequence of any species, not only that of humans. In our hotfoot pursuit of the biology that is unique to humans, we should remember that each species will have winners of their own genome race. Could it be, for example, that accelerated regions in chimpanzees, mice or dogs are linked to the adaptive evolution of their brains, too?

HARs seem to be particularly rare in protein-coding sequences. Instead, they often lie near protein-coding genes that have neurodevelopmental functions, perhaps within regions that are involved in regulating when and where these genes are turned on. Rapid human-specific evolution, and particularly the evolution of brain morphology and of behavioural traits, may thus be associated more with fine-tuning the spatial and temporal expression of protein-coding genes than with altering the molecular functions of their encoded proteins. Previously, the hunt for changes in DNA that are causally linked to human-specific biology had concentrated on differences that would alter the amino-acid make-up of the encoded protein. Now it would seem that searches within the functional non-coding 'dark matter' might be more enlightening. ■

Chris P. Ponting and Gerton Lunter are in the MRC Functional Genetics Unit, University of Oxford, Department of Physiology, Anatomy and Genetics, South Parks Road, Oxford OX1 3QX, UK. e-mail: chris.ponting@anat.ox.ac.uk

CRYSTALLOGRAPHY

Solid oxygen takes shape

Burkhard Militzer and Russell J. Hemley

Oxygen crystallizes into a sequence of structures, starting as an insulator at low pressure and becoming a superconductor at high pressure. The elusive structure of an intermediate phase has now been determined.

Oxygen, the third most abundant element in the cosmos, is fundamental to our world. Its role in the development of chemistry cannot be underestimated. As a character in the play *Oxygen*¹ comments, "...the chemical revolution came from oxygen. The element ought to come first." The events surrounding oxygen's discovery make for lively theatre; its molecular diatomic form (O_2) was identified before chemists understood why two oxygen atoms would bond together at all. It wasn't long before another molecular type was discovered — ozone, later found to be O_3 . Two research groups, Lundegaard *et al.* (page 201 of this issue)² and Fujihisa *et al.* (writing in *Physical Review Letters*)³, have now established that another form of oxygen, best described as $(O_2)_4$, appears under high-pressure conditions.

Oxygen crystallizes on cooling, retaining its diatomic character and forming a unique solid, an elemental molecular magnet; this magnetism arises from the electronic ground state of the O_2 molecule⁴. Under sufficiently high pressure, molecules tend to split into atoms, ultimately resulting in a metallic state. Solid oxygen is not only metallic under pressure, but actually becomes a superconductor near 100 gigapascals⁵ (GPa). At low pressures, solid oxygen is an insulator. With increasing pressure, oxygen crystallizes into a sequence of different crystal structures. The properties of the most intriguing one, which links the low-pressure magnetic and the high-pressure superconducting phases, have finally been determined. This phase is known as ϵ -oxygen, and persists over a broad range of pressures from 8 to 96 GPa at low temperature.

The ϵ -phase of oxygen appears near 10 GPa at room temperature⁶ and shows remarkable optical properties. Crystals of this red solid have striking spectroscopic features that indicate that O_2 molecules persist, but with strong interactions between the molecules⁷. However, the detailed structure of ϵ -oxygen has remained obscure for 27 years, despite numerous attempts to clarify it.

Insight into the properties of ϵ -oxygen has been gained by considering its magnetic state. Magnetism tends to collapse at high pressures in solids. Spectroscopic studies, together with this idea, previously led researchers to propose that O_2 molecules dimerize at high pressures, effectively forming O_4 clusters in which electron spins are paired in a non-magnetic state⁸. However, this hypothesis was not compatible with theoretical studies, which predicted chain-like structures⁹ as the lowest-energy configuration

of atoms for ϵ -oxygen. Recent neutron-diffraction experiments, which directly probed the electronic spin structure of ϵ -oxygen, showed that magnetism indeed collapses as this phase forms¹⁰.

High-pressure diffraction experiments^{2,3}, performed using diamond-anvil-cell techniques, have finally revealed the crystal structure of ϵ -oxygen. Lundegaard *et al.*² use X-ray diffraction from single crystals suspended in helium at selected pressures. In contrast, Fujihisa *et al.*³ measure X-ray diffraction of polycrystalline ϵ -oxygen over a broad range of pressures. Essentially the same structure for ϵ -oxygen is obtained from both studies — a unique arrangement in which diatomic O_2 molecules remain intact, but with additional linkages giving rise to $(O_2)_4$ clusters (Fig. 1). The new structure^{2,3} is also consistent with the observed non-magnetic state of ϵ -oxygen¹⁰.

Why does this structure form? Oxygen is the lightest atom in Group VI of the periodic table, with sulphur being the next heaviest family member. At room pressure and temperature, sulphur crystals form as lattices of S_8 molecules, with the atoms arranged in eight-membered rings. A rule of thumb for high-pressure research is that, under compression, elements of a given family behave like heavier members of the same series. So at first glance, it might not seem surprising that an eight-atom configuration of oxygen atoms could appear under pressure. However, inspection of the ϵ -oxygen structures^{2,3} shows that a simple analogy with S_8 molecules is misleading; the O_2 molecules are arranged in $(O_2)_4$ rhombohedra while preserving their original diatomic character.

So, are the $(O_2)_4$ clusters really a new kind of crystal subunit? And how different is the structure from a uniform packing of molecules that is more common in solids? At 17.6 GPa, the distance between molecules at the edges of the cluster is 2.18 Å, whereas the shortest distance between individual $(O_2)_4$ units is 2.56 Å, consistent with the crystal description in terms of $(O_2)_4$ clusters. This makes the structure unique, and clearly distinguishes it from the previously proposed O_4 clusters or extended chains^{9,11}.

Interestingly, the structures obtained from the different studies^{2,3} are not identical: the single-crystal data² yield a more complete refinement of the atomic positions compared with the powder method³. First-principles calculations demonstrate that a structure derived

- Pollard, K. S. *et al.* *Nature* **443**, 167–172 (2006).
- Semendeferi, K., Lu, A., Schenker, N. & Damasio, H. *Nature Neurosci.* **5**, 272–276 (2002).
- Hill, R. S. & Walsh, C. A. *Nature* **437**, 64–67 (2005).
- King, M. C. & Wilson, A. C. *Science* **188**, 107–116 (1975).
- Waterston, R. H. *et al.* *Nature* **420**, 520–562 (2002).
- Lunter, G. *et al.* *PLoS Comput. Biol.* **2**, e5 (2006).
- Wyckoff, G. J. *et al.* *Nature* **403**, 304–309 (2000).
- Meunier, J. & Duret, L. *Mol. Biol. Evol.* **21**, 984–990 (2004).

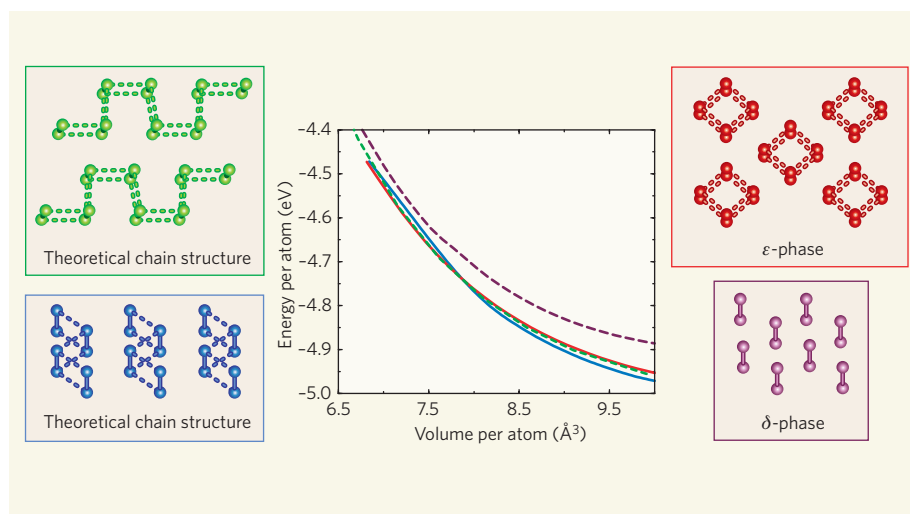


Figure 1 | Structures and energies of dense oxygen. These energy–volume curves for solid oxygen show the energies calculated for different arrangements of molecules at absolute zero. The calculations were performed using density functional theory and the projector augmented-wave method¹⁵. These calculations predict chain structures^{9,11} (blue and green; dotted lines highlight pairs of molecules less than 2.2 Å apart) to be close in energy to the (O₂)₄ structure (red) that has now been observed in the ε-phase of solid oxygen^{2,3}. The new (O₂)₄ structure is not the lowest-energy structure at this level of theory, demonstrating the need for additional studies. Another phase of oxygen, known as the δ-phase (purple), is shown for comparison, as an example of a layered structure⁴.

from single-crystal diffraction has lower energy and is therefore more stable. Such calculations (Fig. 1) also show that the predicted (O₂)₄ structure is very close in energy to the chain-like structures that were previously proposed^{9,11} for ε-oxygen. However, these calculations fail to show that the (O₂)₄ structure has the lowest energy, which is probably why previous theoretical attempts to predict the correct structure for ε-oxygen were unsuccessful. Notably, oxygen in an S₈-like ring structure is significantly higher in energy than all the known oxygen structures.

These results^{2,3} will have broad relevance to other materials that are placed under extreme pressure. For example, as Lundegaard *et al.*² point out, there may be parallels between the behaviour of ε-oxygen and the charge-transfer activity observed in hydrogen at much

higher pressure¹². There is also experimental and theoretical evidence for clustering and formation of polyatomic phases in nitrogen under pressure^{13,14}. The challenge now is to carry out structural and theoretical studies in the higher-pressure metallic and superconducting phases of oxygen. As another character in *Oxygen*¹ opines, without the discovery of oxygen there would be “no chemistry as we now know it”. The findings with ε-oxygen open up a fresh dimension of chemistry that we are only just getting to know.

Burkhard Militzer and Russell J. Hemley are in the Geophysical Laboratory, Carnegie Institution of Washington, 5251 Broad Branch Road NW, Washington, Washington DC 20015, USA. e-mails: militzer@gl.ciw.edu; r.hemley@gl.ciw.edu

1. Djerassi, C. & Hoffmann, R. *Oxygen* (Wiley-VCH, New York, 2001).
2. Lundegaard, L. F., Weck, G., McMahon, M. I., Desgreniers, S. & Loubeyre, P. *Nature* **443**, 201–204 (2006).
3. Fujihisa, H. *et al.* *Phys. Rev. Lett.* **97**, 085503 (2006).
4. Freiman, Y. & Jodl, H. J. *Phys. Rep.* **401**, 1–228 (2001).
5. Shimizu, K., Suhara, K., Ikumo, M., Eremets, M. I. & Amaya, K. *Nature* **393**, 767–769 (1998).
6. Nicol, M., Hirsch, K. R. & Holzapfel, W. B. *Chem. Phys. Lett.* **68**, 49–52 (1979).
7. Goncharov, A. F., Gregoryanz, E., Hemley, R. J. & Mao, H. K. *Phys. Rev. B* **68**, 100102 (2003).
8. Gorelli, F. A., Ulivi, L., Santoro, M. & Bini, R. *Phys. Rev. Lett.* **83**, 4093–4096 (1999).
9. Neaton, J. B. & Ashcroft, N. W. *Phys. Rev. Lett.* **88**, 205503 (2002).
10. Goncharenko, I. N. *Phys. Rev. Lett.* **94**, 205701 (2005).
11. Oganov, A. R. & Glass, C. W. *J. Chem. Phys.* **124**, 244704 (2006).
12. Hemley, R. J., Soos, Z. G., Hanfland, M. & Mao, H. K. *Nature* **369**, 384–387 (1994).
13. Gregoryanz, E. *et al.* *Phys. Rev. B* **66**, 224108 (2002).
14. Mattson, W. D., Sanchez-Portal, D., Chiesa, S. & Martin, R. M. *Phys. Rev. Lett.* **93**, 125501 (2004).
15. Kresse, G. & Furthmüller, J. *Comput. Mater. Sci.* **6**, 15–50 (1996).

ASTRONOMY

Dawn after the dark age

Richard McMahon

The latest surveys provide evidence for one, maybe two, galaxies farther back in cosmic time than ever detected before. But does the fact that we don't see more mean these are the very first galaxies to be formed?

Determining when the first stars and galaxies formed is a matter of profound importance: fuelled by primordial hydrogen, these bodies triggered the nucleosynthesis of the heavier elements, such as carbon, nitrogen and oxygen, that are the basis of life. By studying the first galaxies, we can also hope to understand how the Universe formed and evolved, and detect the younger progenitors of galaxies like our own Milky Way. In this issue, Iye *et al.* (page 186)¹ report the discovery of the most distant galaxy yet, one whose photons must have left it 12.7 billion years ago, when the Universe was just 750 million years old, and some 8 billion years before the Sun and Earth were formed. And Bouwens and Illingworth

(page 189)² report on their search for galaxies even farther away, more than 13 billion light years from Earth.

The study of the most distant galaxies has much in common with archaeology: the farther back one looks, the scantier the evidence becomes, and the harder it is to draw conclusions. Two factors are responsible. First, there are the immense distances of around 10²⁴ km implied by light journey times of more than 10 billion years; the brightness of a source diminishes with the inverse square of its distance. Second, there is the expansion of the Universe, which stretches the wavelength of light from distant objects by a factor 1 + z. The quantity z is known as the redshift, as the expansion

moves all observed wavelengths towards longer, redder wavelengths. The older an object is, the greater its redshift; but unfortunately, the redder one gets, the brighter the night sky becomes. As a result, searching for the very earliest objects becomes — from the ground, at least — increasingly difficult.

Like archaeology, astronomy also has its Dark Ages from which evidence is particularly sparse. As the Universe expanded and cooled, neutral hydrogen and helium were created from the hot plasma of matter at the so-called epoch of recombination, around 400,000 years after the Big Bang. As very few atoms remained ionized in this neutral Universe, very little radiation can be detected from this era. The

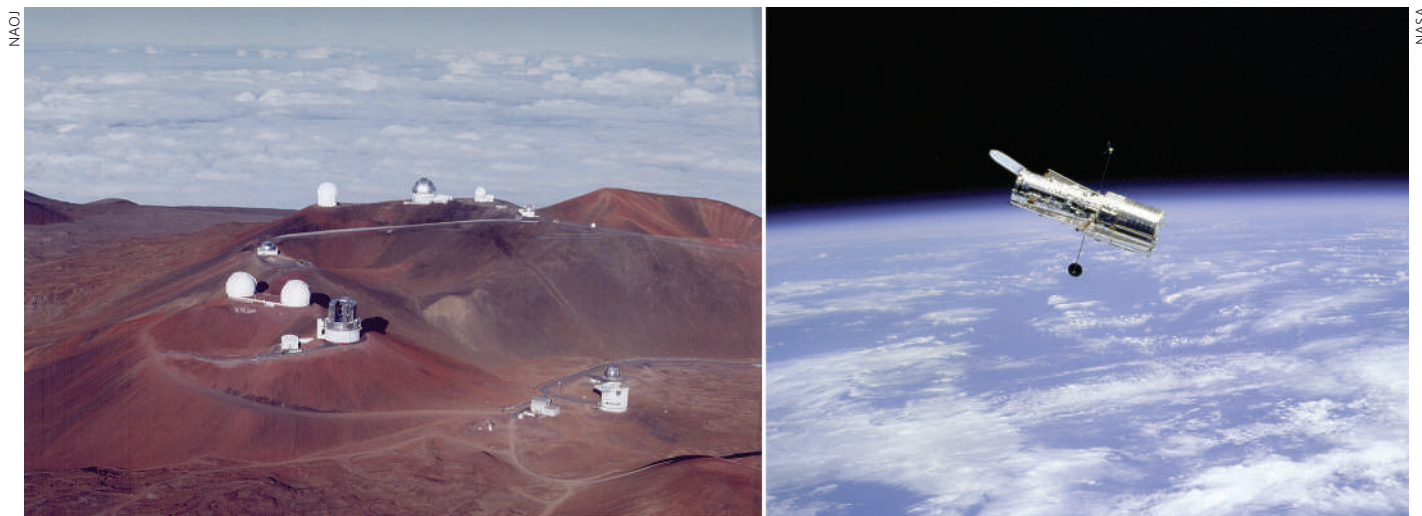


Figure 1 | On Earth as it is in heaven. The Subaru and Hubble Space telescopes, used by Iye *et al.*¹ and by Bouwens and Illingworth² in their respective studies of early galaxies, overlook the clouds — the one some 4 km up on the summit of Mauna Kea on Hawaii, the other from orbit at an altitude of around 600 km over Earth's surface.

situation changed only when the first generation of luminous sources — massive stars, galaxies and accreting black holes — reionized and 'lit up' the gas in the Universe.

In recent years, the search for galaxies has been pushed back progressively towards these trailblazers. Improvements in solid-state detector technology have followed Moore's law, with a doubling of chip density every two years or so; mega-pixel devices and highly efficient, large-format red-sensitive detectors have arrived on wide-field imaging instruments, on a new generation of 8-metre-aperture telescopes, and on the refurbished Hubble Space Telescope (HST). From a redshift of 4.55 in 1996 (ref. 3), the earliest star-forming galaxy had been put back to $z = 6.56$ by 2002 (ref. 4).

Iye and colleagues' discovery¹ takes us farther back to $z = 6.96$. But the authors note that, by comparison with observations at smaller redshifts, they would have expected to find around five galaxies in a survey of their scale. Bouwens and Illingworth² recount a similar story at redshifts of 7–8. Where they might by extrapolation have expected to find around ten galaxies, they found only one unconfirmed candidate. So could we now really be looking back to the very earliest phase of galaxy formation at the epoch of reionization? With the current results it is still hard to tell.

The two surveys^{1,2} employed quite different experimental approaches. Iye *et al.*¹ used the 8.2-m-aperture Subaru telescope on Hawaii (Fig. 1), one of the largest ground-based telescopes, together a mosaic of CCD detectors with a total of 84 million pixels. The authors exploited a gap in the intense hydroxyl airglow radiation that is created by processes high in Earth's atmosphere and dominates the terrestrial night sky at wavelengths above 700 nm. Using a special interference filter centred on 975.5 nm with a bandwidth of 20 nm, they could search between hydroxyl lines for so-called Lyman- α radiation. This

unique, asymmetric spectral signature is typical of star formation, and comes from photons produced by a ground-state transition in hydrogen. In this case¹, the focus was on an ultraviolet wavelength of 121.6 nm. At redshift 6.96, the observed wavelength is stretched by a factor of 7.96. The resulting radiation at 968.2 nm is beyond the range of the human eye, and almost at the limit of conventional silicon-based CCD detectors. Even so, we can expect this ground-based technique to be extended to longer wavelengths in the coming years⁵.

Bouwens and Illingworth's identification method² used the 2.5-m-aperture HST (Fig. 1). As this is above Earth's atmosphere, it is not affected by airglow radiation effects and can work at longer wavelengths (and so higher redshift). The authors looked for a break in the radiation between the infrared and the optical regions that is caused by absorption through neutral hydrogen along the line of sight, and found this in one case. Unlike the characteristic Lyman- α emission, this break is not an unambiguous signal of a high-redshift galaxy with high rates of star formation, but could have been caused by interstellar dust or intrinsic stellar features in galaxies at lower redshifts. For this reason, independent spectroscopic confirmation is required for the existence of this galaxy. As the HST does not carry appropriate instrumentation, this might have to await the launch, currently projected for 2013, of Hubble's successor, the 6-m-aperture James Webb Space Telescope.

Although they could search at longer wavelengths, the smaller aperture and older detectors on the HST meant that Bouwens and Illingworth's observations² were confined to a patch of sky about 50 times smaller than that covered by Iye and colleagues' optical detectors¹. Whereas the Subaru telescope needed just a few nights to collect the requisite data, the HST needed a few years. That also limits the scope for extending the HST work to larger

areas of sky within the few remaining years of the telescope's projected lifetime. In the shorter term, Iye and colleagues plan to search another patch of sky to increase their sample size. Only then will the presence, or absence, of further galaxies be able to tell us whether we really are homing in on the era of reionization.

Other techniques exist to search for galaxies at higher redshifts. A candidate galaxy at $z = 7$ has been reported using gravitational lensing — the bending of light by massive intervening bodies — to amplify images of distant objects that would otherwise be undetectable⁶, although airglow has prevented spectroscopic confirmation of this galaxy, too. Indirect evidence for galaxies at still higher redshifts comes from recent observations by NASA's WMAP satellite of polarization in the cosmic microwave background radiation, which was left behind at the epoch of recombination. A galaxy spotted at $z = 6$ might have formed its first stars at up to $z = 13.5$ — just 300 million years after the Big Bang⁷.

The success of Iye *et al.*¹ bodes well for continued searches for the first galaxies to emerge from the Universe's dark ages. Bouwens and Illingworth's conclusions² are no less important, reminding us just how difficult this task might be. Armed with both sets of results, we can design better experiments to increase the probability that still older galaxies can be discovered.

Richard McMahon is at the Institute of Astronomy, University of Cambridge, Cambridge CB3 0HA, UK.
e-mail: rgm@ast.cam.ac.uk

1. Iye, M. *et al.* *Nature* **443**, 186–188 (2006).
2. Bouwens, R. J. & Illingworth, G. D. *Nature* **443**, 189–192 (2006).
3. Hu, E. M. & McMahon, R. G. *Nature* **382**, 231–233 (1996).
4. Hu, E. M. *et al.* *Astrophys. J.* **568**, L75–L79 (2002).
5. Horton, A. *et al.* *Proc. SPIE* **5492**, 1022–1032 (2004).
6. Kneib, J.-P., Ellis, R. S., Santos, M. R. & Richard, J. *Astrophys. J.* **607**, 697–703 (2004).
7. Eyles, L. *et al.* *Mon. Not. R. Astron. Soc.* **346**, 443–454 (2005).

CANCER BIOLOGY

Can less be more for p53?

Anton Berns

The tumour-suppressor protein p53 is often mutated in cancer. But it seems that p53 deficiency is not all bad, and inhibiting this protein might mitigate toxic side effects of chemotherapy and radiotherapy.

The p53 protein gained its reputation as the 'guardian of the genome' because it mediates the cell's response to certain types of injury, relaying signals from both the DNA-damage response pathway and the oncogenic (tumour-promoting) stress pathway. Defects in these pathways are associated with a predisposition to cancer, and because they converge on p53, it has always been assumed that they act primarily through this crucial protein. However, this view might need a rethink in light of work from Christophorou *et al.* (page 214 of this issue)¹ and Efeyan *et al.* (page 159)².

To dissect the relative contributions of signals that lead to the activation of p53, Christophorou *et al.* used genetically engineered mice called p53ER^{TAM} mice. In these animals, p53 production is disabled during development, but it can be restored later in life because the gene encoding p53 can be activated by the drug tamoxifen³. In normal, 'wild-type' mice, when radiation damages their DNA, the cells undergo programmed cell death (apoptosis). However, in mice that have no p53 — either mice where the p53 gene has been permanently inactivated (p53-null mice) or p53ER^{TAM} mice without tamoxifen — this almost instantaneous response does not occur after irradiation. Mice that lack p53 are also prone to tumours.

The authors then examined whether temporary restoration of p53 function would influence the radiotoxicity response and tumour development, and if so, whether p53 exerts its effects at the time of DNA damage or at some point afterwards. To address this, they restored p53 expression for 6 days, either directly before whole-body radiation (2.5 gray) or 8 days after radiation. When p53 function was restored during the 6 days before radiation, a pronounced apoptotic response was seen in all radiosensitive tissues, similar to that observed in wild-type mice. Surprisingly, despite this response, tumour onset was not delayed. Thus, the apoptotic response failed to eliminate incipient tumour cells. By contrast, when p53 function was restored

8 days after DNA damage, there was no apoptotic response to radiation, but tumour onset was markedly delayed (Fig. 1).

These results suggest that the presence of p53 at the time of DNA damage is not strictly required for tumour suppression. Indeed, the authors argue that p53 exerts its tumour-suppressive function primarily through the oncogenic-stress pathway rather than the DNA-damage pathway, at least in this experimental system. In support of this notion, the authors observed that the tumour-suppressive effect of delayed p53 restoration was lost when the mice also lacked a principal factor in the oncogenic-stress pathway called p19Arf. This suggests that p19Arf is crucial in relaying the tumour-suppressive signal. Furthermore, tumours emerged significantly earlier in double-mutant mice lacking both p19Arf and p53

(that is, p19Arf^{-/-}; p53ER^{TAM} mice) than in the p53ER^{TAM} mice without a functional p53. This observation supports results⁴ indicating that p19Arf has tumour-suppressive functions that are independent of p53.

Efeyan *et al.*² confirm that p19Arf is crucial in protecting against tumours. They engineered mice that have an extra copy of the gene encoding p53 (p53^{super} mice), and which have greater resistance to certain chemically induced cancers (lymphomas, sarcomas and haemangiosarcomas). But this extra protection was not observed when p19Arf was disabled in the p53^{super} mice.

These are provocative observations. Notably, following DNA damage, the cells killed by p53-mediated apoptosis are apparently irrelevant to tumour development in this model. They have either lost the capacity to give rise to tumours as a result of the DNA damage incurred, or they lack tumour-inducing capacity altogether. One has to assume, then, that incipient tumour cells arise from a distinct cell population that wards off this DNA-damage-induced, p53-mediated cell death. In these cells, the higher mutation load from the radiation is probably responsible for the accelerated tumour onset. The oncogenic mutations induced by the radiation will upregulate p19Arf. But in this model, by the time this can happen, p53 responsiveness

will have faded because of the withdrawal of tamoxifen. Consequently, early restoration of p53 function has no beneficial effect whereas delayed restoration permits p19Arf to activate p53. This leads to elimination or permanent arrest of any incipient tumour cells, and postpones tumour onset.

These results underscore the value of the oncogenic-stress pathway in preventing tumorigenesis as well as the capacity of p53 to impair tumorigenesis, even if it is functional for only a short time. However, the transient functionality of p53 cannot prevent the defects in the cell-division cycle and abnormal chromosome numbers that are the hallmarks of p53-mutant tumours. Also, it cannot substitute for the protective surveillance through the DNA-damage pathway that might effectively cope with mild DNA damage⁵.

Will p53 and p19Arf act in the same way in other tumour types? In both papers^{1,2}, the tumour-suppressive capacity of p53 seems to be completely dependent on the presence of p19Arf, but in other studies, loss of p19Arf had almost no effect on p53-mediated tumour suppression⁶. This suggests that the effects of p53 are context dependent and that different signalling

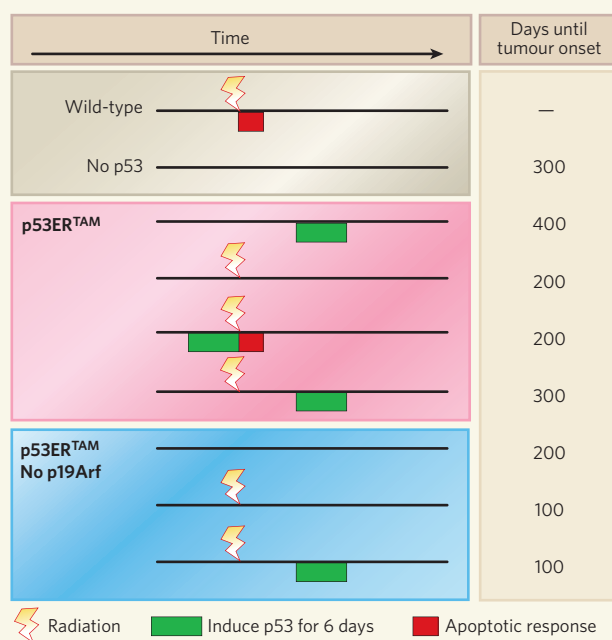


Figure 1 | Effects of temporary restoration of p53 protein on radiotoxicity and tumour latency. Christophorou *et al.*¹ engineered a mouse strain called p53ER^{TAM} that lacks p53 but in which expression of the p53 gene can be restored for short periods. The timing of p53 restoration relative to irradiation determines whether massive apoptosis (cell death) occurs without a delay in tumour onset, or whether tumour development is delayed. The protection from tumour development conferred by 6 days of p53 restoration is seen in both irradiated and non-irradiated mice. The absence of a delay in tumour development in irradiated mice in which p53 was restored early indicates that tumours in irradiated mice arise as a result of radiation exposure, rather than from pre-existing lesions. (Data were compiled from Christophorou *et al.*^{1,3} and G. Evan, personal communication.)

routes can be used in tumour suppression⁷.

The studies^{1,2} point to a potential clinical application. Counterintuitively, because tumorigenesis is not delayed by the extensive cell death seen after radiation in the absence of p53, temporary inactivation of p53 during radiation or certain drug treatments that induce DNA damage might be a relatively safe approach to protect normal cells from death, thus permitting a higher dose of these cell-killing agents. This could be particularly useful for the treatment of tumours in which p53 has been mutated or deleted. Support for this strategy comes from the observation that high-dose chemotherapy seems particularly effective in p53-deficient cancers⁸. In addition, a p53-deficient mouse strain called FVB shows a decreased apoptotic response following radiation⁹, suggesting that the inactivation of p53 throughout the body can permit a higher dose of radiation or of DNA-damaging drugs.

Therefore, it will be vital to analyse a range of mouse model systems, to make sure that inactivating p53 at the same time as inducing DNA damage with radiation or drugs does not promote new tumours. But if it does not, the therapeutic avenue of inhibiting p53 concomitantly with the administration of DNA-damaging drugs or radiation is worth further exploration. The p53 status of a tumour would then become a critical parameter for deciding on treatment. In p53-containing tumours, p53

might be exploited by enhancing apoptosis¹⁰ with low doses of DNA-damaging agents or radiation therapy. In p53-deficient tumours, the perturbation of the cell cycle by p53 loss would be used to drive tumour cells to cell death resulting from abnormal cell division, while protecting normal cells from p53-mediated apoptosis by systemic inhibition of p53. Inhibitors of p53 such as pifithrin¹¹ could be used to test this hypothesis. Who would have predicted that disabling p53 systemically would one day be considered as a promising treatment option? ■

Anton Berns is in the Division of Molecular Genetics and Centre of Biomedical Genetics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands. e-mail: a.berns@nki.nl.

1. Christophorou, M. A., Ringhausen, I., Finch, A. J., Brown Swigart, L. & Evan, G. I. *Nature* **443**, 214–217 (2006).
2. Efeyan, A., Garcia-Cao, I., Herranz, D., Vaelasco-Miguel, S. & Serrano, M. *Nature* **443**, 159 (2006).
3. Christophorou, M. A. *et al.* *Nature Genet.* **37**, 718–726 (2005).
4. Sherr, C. J. *et al.* *Cold Spring Harb. Symp. Quant. Biol.* **70**, 129–137 (2005).
5. Bensaad, K. & Vousden, K. H. *Nature Med.* **11**, 1278–1279 (2005).
6. Tolbert, D., Lu, X., Yin, C., Tantama, M. & Van Dyke, T. *Mol. Cell. Biol.* **22**, 370–377 (2002).
7. Eymir, B. *et al.* *Mol. Cell. Biol.* **26**, 4339–4350 (2006).
8. Bertheau, P. *et al.* *Lancet* **360**, 852–854 (2002).
9. Westphal, C. H. *et al.* *Cancer Res.* **58**, 5637–5639 (1998).
10. Schmitt, C. A. *et al.* *Cancer Cell* **1**, 289–298 (2002).
11. Strom, E. *et al.* *Nature Chem. Biol.* **2**, 474–479 (2006).

of research in which the imprint of quantum effects is beginning to be seen on a macroscopic scale^{5,6}. It might seem odd to describe these objects, which have dimensions of less than a micrometre, as macroscopic, but the crucial point is that such systems typically contain billions of constituent atoms. And yet, at low temperatures, rather than behaving classically, these large atomic ensembles can begin collectively to manifest quantum behaviour of the type commonly associated with individual atoms and molecules. The field known as quantum electromechanics focuses on such coherent mechanical systems, which are poised midway between two seemingly antithetical domains.

Naik and colleagues³ explore the interactions between a cryogenically cooled nanomechanical resonator and a nanoscale device called a superconducting single-electron transistor (SSET; Fig. 1). The resonator itself is an object somewhat like a guitar string that, when excited, vibrates at radio frequencies (about 21 megahertz) and amplitudes typically of only atomic dimensions. The crucial step to enable observations on the system is to transduce these fast, tiny vibrations — that is, to convert them into electrical signals.

The authors' readout uses the parallel development of high-frequency motion detectors based on the conventional (non-superconducting) single-electron transistor⁷, and their own previous work⁸ with the SSET. Over a measurement time of a second, they achieve a record sensitivity with the SSET of just a fraction of a femtometre (a femtometre is 10^{-15} m). This is about the size of a helium nucleus, and much smaller than an atom.

An intriguing aspect of the resulting measurements³ is the detail they reveal about the noise processes that limit the sensitivity of the measurements themselves. A tight coupling between the SSET and the resonator is the key to precise measurements, and this is achieved by creating a charge interaction between the devices. When the authors tune this electrostatic interaction to maximize the detection sensitivity, the intrinsic random fluctuations of charges within the SSET (the observer), which are an inextricable consequence of its operation, begin to affect the nanomechanical resonator (the observed). This effect is called back-action, and is predicted — in fact, guaranteed — by quantum mechanics, irrespective of how a measurement is made.

In the case in hand, as the SSET 'observes' the resonator, its charge carriers tunnel quantum mechanically to and from an isolated, nanoscale conducting island within it. These charge fluctuations produce a stochastic back-action force that randomly drives the resonator. Charge is discrete, and the associated 'granularity' of these fluctuations, and their impact on the nanomechanical resonator, reveal the underlying nature of the SSET's fundamental particles — electrons in various correlated quantum-mechanical states, known as quasiparticles and Cooper pairs.

QUANTUM PHYSICS

Observing and the observed

Michael Roukes

Quantum mechanics states that the measurement process can fundamentally alter what is being measured. This 'back-action' has been observed on the macroscopic scale — in the vibrations of a tiny mechanical device.

In daily life, we continuously observe and then assimilate these observations. The cumulative effect of this process grounds our intuition; our intuition, in turn, governs our next actions. But, oddly, our deep sense of the order of things acquired in this way does not square with fundamental tenets of quantum mechanics. Many people have found this discrepancy profoundly disconcerting: Einstein, for instance, famously decried one of the least intuitive consequences of quantum theory as "spooky action at a distance"^{1,2}. In this issue, Naik *et al.* (page 193)³ provide a beautiful example of how nanomechanical systems can be used to explore such quandaries. The authors induce cooling of a tiny mechanical resonator just by measuring its vibrations. Closer inspection shows that the phenomenon is a mechanical analogue of the familiar optical cooling of trapped atoms.

Quantum mechanics is about both discreteness and discreteness. On the first property, the

theory states that there are fundamental limits to how precisely we can 'know' a system, and that these limits are dictated by, for example, the quantization of charge and energy. Regarding the second property, quantum mechanics says that no free 'sneak peeks' are allowed: observation always comes at the price of interaction, and this interaction will irrevocably change the fate of the object observed. (Actually, a special class of sneak peaks is possible; see, for example, ref. 4.)

For the macroscopic things of our everyday world, the consequences of these interactions generally turn out to be negligible. But at the microscopic level, they govern how the fundamental building-blocks of our world interact. The disparity between the macroscopic and microscopic realms has intrigued physicists, and fuelled heated discussions, since quantum mechanics first emerged in the 1920s.

Nanomechanical systems are one area

ENTOMOLOGY

To catch a bee

Many organisms rely on other species to transport them from one spot to another — particularly in harsh environments where such meagre resources as there are tend to occur in clumps. As Leslie S. Saul-Gershenz and Jocelyn G. Millar report (*Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0603901103; 2006), one species of the blister beetle *Meloe franciscanus* is a particularly innovative passenger: it can hail its taxi ride.

This blister beetle lives in the deserts of the southwestern United States. It feeds and lays its eggs under a plant that also provides nectar for the beetles' host and transporter species, a solitary bee of the species *Habropoda pallida*.

Larvae of the beetle cooperatively form a spherical mass on the plant (right image), and simply hitch a ride when a male bee intent on mating with a female makes contact (left image). When the infested male copulates with a real female, the larvae are transferred and carried to the bee's nest. There they set up camp, and complete their development into adults nourished by the pollen and nectar stores of the nest and by the bee's egg.

Saul-Gershenz and Millar set out to test what lures the bees to make contact with the larvae in the first place. They found that visual models of the larval aggregations held no interest for the bees — but if the models were scented with



an organic extract from either the larvae or the female bee's head, the bees found them just as enticing as real larvae.

Comparison of chemical profiles of the larval and bee extracts identified two alkene molecules common to the larvae and female bees that were



not present in the males. And this heady blend of alkenes did indeed attract the male bees. It seems the beetle larvae have evolved a way to exploit the bee's sexual communication system as a means of calling a cab.

Helen Dell

Quantum electromechanics is still an emerging field. In experiments with nanomechanical devices, making the transition from the classical realm to the quantum world means that the random thermal fluctuations must be frozen out, allowing the so-called zero-point motion to come to the fore. This is achieved by cooling the devices to such low temperatures that the occupation of the fundamental vibrational energy modes is suppressed. But it is precisely in this cryogenic regime that thermal contact between these modes and the outside world

falls precipitously⁹, making it increasingly difficult to cool a device into the quantum regime.

An unexpected consequence of the latest work³ may help in this matter. Naik and colleagues show that orchestrating a strong coupling between the cooled SSET and the nanomechanical resonator actually chills the latter. In other words, the very act of observing can make an observed object colder. Closer analysis shows that this surprising effect is formally analogous to the phenomenon of 'optical' trapping and cooling of atoms by

laser light^{10–12}. Both effects rely on back-action from an ensemble of discrete particles — whether correlated electrons or laser photons. These particles interact resonantly at a rate that depends on the motion of the observed object (the nanoresonator or atoms to be cooled).

In both cases, the back-action forces are engineered to damp, rather than excite, the object's motion. The overall effect is that the fluctuating interactions with the object provide it with a 'thermal bath', at a temperature lower than ambient, into which its energy can be absorbed. The puddle of barely moving atoms that resulted from optical cooling and trapping earned the name 'optical molasses'. In recognition of the inspiration drawn from that work, which won the 1997 Nobel physics prize, the authors call their effect 'Cooper-pair molasses'. From here, further advances along these lines will engender even deeper penetration into the strange realm of quantum electromechanics, which lies at the junction between the classical and quantum worlds.

Michael Roukes is in the Departments of Physics, Applied Physics and Bioengineering and at the Kavli Nanoscience Institute, California Institute of Technology, Pasadena, California 91125, USA. e-mail: roukes@caltech.edu

1. www.physics.mcgill.ca/~burkes/Bell/Bell.html
2. www.upscale.utoronto.ca/PVB/Harrison/BellsTheorem/BellsTheorem.html
3. Naik, A. et al. *Nature* **443**, 193–196 (2006).
4. Braginsky, V. B. & Khalili, F. Ya. *Rev. Mod. Phys.* **68**, 1–11 (1996).
5. Cho, A. *Science* **299**, 36–37 (2003).
6. Schwab, K. C. & Roukes, M. L. *Phys. Today* **58** (7), 36–42 (2005).
7. Knobel, R. & Cleland, A. N. *Nature* **424**, 291–293 (2003).
8. LaHaye, M. D. et al. *Science* **304**, 74–77 (2004).
9. Schwab, K. C., Henriksen, E. A., Worlock, J. M. & Roukes, M. L. *Nature* **404**, 974–977 (2000).
10. Chu, S. *Rev. Mod. Phys.* **70**, 685–706 (1998).
11. Cohen-Tannoudji, C. N. *Rev. Mod. Phys.* **70**, 707–719 (1998).
12. Phillips, W. D. *Rev. Mod. Phys.* **70**, 721–741 (1998).

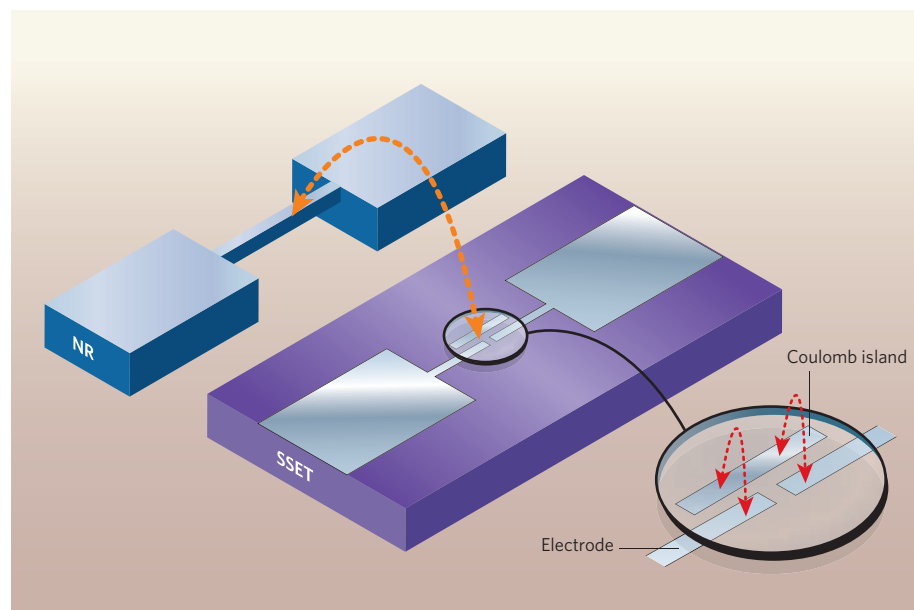


Figure 1 | Cooper-pair molasses in action. Naik and colleagues' experimental set-up³ consists of a nanomechanical resonator (NR) comprising a suspended, doubly clamped, nanometre-scale beam electrostatically coupled (orange arrow) to a superconducting single-electron transistor (SSET). Back-action arises from the intrinsic charge fluctuations during the operation of the device. These fluctuations are due to discrete quantum-mechanical tunnelling events of quasiparticles or Cooper pairs (red arrows) between the electrodes and an isolated, conducting 'Coulomb island'.



50 YEARS AGO

'World Population Growth and its Regulation by Natural Means' — Recent discussions on the future of humanity have tended to be rather discouraging, and there are plenty of eminent authorities who are prophesying disaster. World population is now multiplying so fast that at the present rate it is bound eventually to outgrow its food supply, however much that can be increased, and then famine and disease will intervene to prevent further population growth, unless indeed the hydrogen bomb provides an effective and relatively humane alternative to more orthodox Malthusian agents. The only solution seems to be the immediate and universal application of a policy of deliberate family limitation, and few are so optimistic as to believe that this is either practicable or likely to be effective in time to save us from the crash.

From *Nature* 15 September 1956.

100 YEARS AGO

The September issue of the new bi-monthly journal *Concrete* contains admirably illustrated articles on the micro-structure of Portland cement by Dr. C. H. Desch and on reinforced concrete at the Milan Exhibition by Mr. F. R. Farrow. This new addition to the technical periodical literature should prove a valuable source of information to all workers in concrete and cement. The details of the new uses to which concrete and reinforced concrete are put are very remarkable. The use of reinforced concrete as a substitute for timber in exposed positions is rapidly increasing. Railway sleepers, telegraph posts, and fence posts are being tried, and efforts are being made to prove that reinforced concrete is an excellent substitute for brick-work where structures of great height are required.

Also:

We notice with deep regret the announcement that Prof. Ludwig Boltzmann, professor of theoretical physics at the University of Vienna, died by his own hand at Duino a few days ago.

From *Nature* 13 September 1906.

STRUCTURAL BIOLOGY

The ins and outs of drug transport

Shimon Schuldiner

Multidrug transporters provide cells with a defence against toxic chemicals, but they are also responsible for drug resistance. The structures of two such transporters reveal novel aspects of their mechanisms.

Living organisms are constantly assailed by a host of harmful chemicals from the environment. Because of the diversity of these 'xenobiotics', cellular survival mechanisms must deal with an immense variety of molecules. Multidrug transporters (MDTs) supply one such strategy. These ubiquitous membrane proteins recognize a wide array of compounds and remove them from the cell in an energy-dependent manner. As a consequence, the concentration of the offending compound is lowered within the cell, so the compound loses its effectiveness against its cellular target and the cell becomes resistant to it. Such a survival strategy for the individual organism may pose serious problems for the treatment of infectious disease or cancer with drugs.

MDTs have evolved into many different forms to act on a wide range of xenobiotics. The structures of two MDTs^{1,2} give an unprecedented view of their mode of action. Murakami *et al.* (page 173 of this issue)¹ and Seeger *et al.* (writing in *Science*)² report the structure of AcrB, a transporter that uses the energy of the

transmembrane proton gradient. And Dawson and Locher (page 180 of this issue)³ detail the structure of Sav1866, a transporter that derives its power from the breakdown of ATP molecules. Despite their different structures and energy sources, the transport mechanisms of the two proteins have a common feature: an inward-facing conformation with the substrate-binding site accessible from the cell interior, and an outward-facing conformation with an extrusion pocket exposed to the external medium. The transition between the two conformations is energy dependent.

AcrB is a major MDT from *Escherichia coli* that has many close relatives in other pathogenic bacteria and that confers drug tolerance to these bacteria. AcrB functions in a complex with two accessory proteins — TolC and AcrA — to export a wide variety of substrates. AcrB determines the substrate specificity of the complex, and is also the site where its energy is used. The structures reported by Murakami *et al.*¹ and Seeger *et al.*² are consistent with previous structures^{4,5} in that they show a threefold

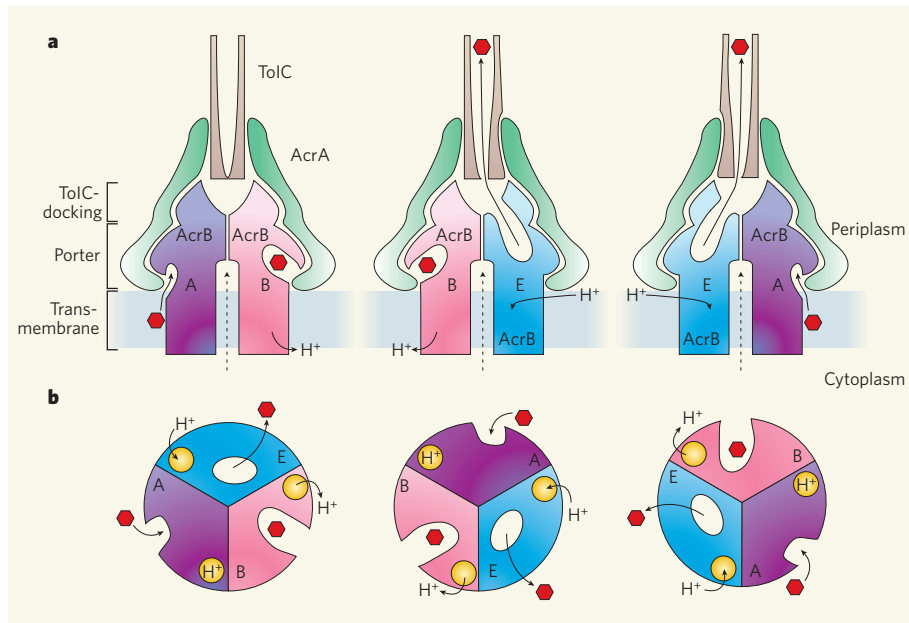


Figure 1 | The structure of the AcrB-drug complex and proposed mechanism of drug transport^{1,2}.

a, The complex is seen from the side, with the drug shown as a red hexagon. The dotted line indicates a possible pathway for substrates moving from the cytoplasm. In this context, Yu *et al.*⁵ propose other locations for drug binding sites to the ones shown here. The complex contains three molecules of AcrB, AcrA accessory proteins and the TolC channel to the exterior. The drug is proposed to enter AcrB when it is in the access (A) conformation, before binding more closely to the porter domain of AcrB in the binding (B) conformation. It is then transported to the opposite face and is released from the extrusion (E) conformation of AcrB. Transport of the xenobiotic is powered by the proton (H⁺) gradient across the membrane. **b**, The proposed ordered multidrug binding change mechanism of the three-unit AcrB complex. (Adapted from Murakami *et al.*¹ and Seeger *et al.*².)

symmetry with three large subdomains. The functional unit is composed of three identical AcrB units (protomers), each with a transmembrane domain, a so-called porter domain, and a TolC-docking domain (Fig. 1a). The porter domain faces the periplasmic space between the plasma membrane and the outer membrane, where the substrate enters into the substrate-binding pockets. The TolC-docking domain has a central funnel that collects the substrate from each protomer and delivers it to the TolC protein located in the outer membrane.

The new AcrB structures show a noteworthy deviation from symmetry that could not be detected in the previous structures: in the AcrB–drug complex, each of the three protomers has a different conformation. The bound substrate is seen in only one of the three protomers (the ‘binding’ protomer; Fig. 1b). Exit from one of the vacant binding sites is open towards the TolC funnel, suggesting that this protomer is the form present just after extrusion of the substrate (the ‘extrusion’ protomer). The other vacant binding site looks to be the state just before substrate binding (the ‘access’ protomer).

What causes this asymmetry and how it is connected to proton movement (the energy source) remains unclear. However, hints are provided by differences between the three protomers in the transmembrane region where three amino-acid residues were previously suggested to be involved in proton transport. In the ‘access’ and the ‘binding’ protomers, a positively charged lysine (Lys 940) is connected by salt bridges to two aspartic acids (Asp 407 and Asp 408). By contrast, in the ‘extrusion’ protomer, Lys 940 is turned nearly 45° towards a threonine (Thr 978) in the transmembrane region TM11 of the protein, and the salt bridges are abolished. This change may cause the twisting of two transmembrane helices that would influence the movement of other subdomains, including those in the porter domain.

The structure suggests a mechanism whereby drugs enter the binding domain, move through the uptake channel and bind in the large internal pocket. At this point, exit from the binding site is blocked. In the extrusion state, however, the entrance closes and the exit opens. The bound drug is pushed out into the funnel by distortion of the binding pocket. These changes are probably coupled to proton movement across the membrane, and may be induced by protonation and deprotonation of Lys 940, Asp 407 and Asp 408 in the transmembrane domains.

The structure reported by Dawson and Locher³ is of Sav1866, an MDT from *Staphylococcus aureus*. It is a member of the large ‘ABC transporter’ family and, as expected from previous studies of this family (for a review see ref. 6), its basic architecture consists of two identical units, with two transmembrane domains that provide a pathway across the membrane, and two cytoplasmic nucleotide-binding domains (NBDs) that break down the ATP nucleotide to provide energy. The structure shows the ATP-

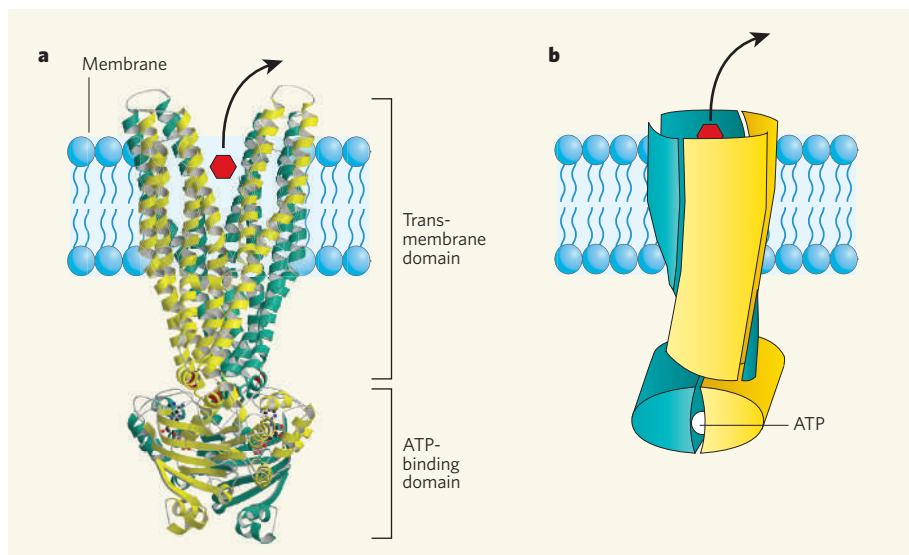


Figure 2 | The structure of the Sav1866 transporter in the ATP-bound conformation³. **a**, Backbone of the Sav1866 transporter, carrying a drug (red), in ribbon representation. The two subunits are shown in yellow and turquoise. **b**, The interaction between the membrane domains and the substrate pathway. (Adapted from Dawson and Locher³.)

bound state, where the NBDs tightly sandwich the nucleotide between them (Fig. 2a). This is consistent with many kinetic and biochemical studies and with structures of isolated NBD domains containing trapped ATP (for example, from the archaeal ABC protein MJ0796)⁷. It differs, however, from two other ABC-transporter structures: that of the *E. coli* vitamin B12 transporter BtuCD crystallized in the absence of nucleotide⁸, and that of the *Staphylococcus typhimurium* transporter called MsbA (refs 9, 10).

Probably the most surprising feature of the new structure³ is that, in the middle of the membrane, bundles of transmembrane helices diverge into two discrete ‘wings’ that point away from one another and consist of helices TM1–TM2 from one subunit and TM3–TM6 from the other. So, rather than being aligned side-by-side, the two subunits of Sav1866 are intricately interleaved. Given these constraints, they are unlikely to move independently and their maximum separation during the reaction cycle is therefore limited (Fig. 2b). This view contradicts mechanistic models that suggest that the NBDs join together upon binding ATP and then dissociate upon completing a transport cycle.

In both structures, the power provided by the energy release (either from the proton gradient or ATP) results in long-range modifications in the accessibility of the xenobiotic-binding site(s) that are transmitted by twisting and tilting of one or more transmembrane helices. In the case of Sav1866, the tight association between the two transmembrane domains was unexpected. In the case of AcrB, the surprising asymmetric structure seems to be essential for function and dictates the ordered occupancy of the binding site. This ordered binding mechanism, termed functional rotation by Murakami *et al.*, is similar in principle to the ATP breakdown/synthesis mechanism of the enzyme F₁F₀-ATPase, except that in AcrB there is no mechanical rotation. It

is possible that this is a remnant of the evolutionary process that led to the development of true rotary molecular machines.

The AcrB structure also provides a neat explanation of how the cell increases the effectiveness of the system, because one TolC molecule is sufficient to eliminate xenobiotic substrates from three monomers that may be removing three different chemical species. How the substrates are removed from the cytoplasm is not yet clear, however. There is compelling evidence that AcrB eliminates many substrates directly from the periplasmic space before they even enter the cell, but it is very likely that it can also expel xenobiotics that have entered the cytoplasm.

Extensive studies will be needed to understand the transport mechanism fully, but the new structures^{1–3} provide some of the ins and outs of the overall process. In addition, the work suggests that a careful re-evaluation of previous structures^{4,5,8–10} is called for. The structures may prove to be essential tools in tackling the serious problems posed by drug resistance and the diseases caused by mutations in the human versions of these transporters.

Shimon Schuldiner is in the Department of Biological Chemistry, Institute of Life Sciences, Hebrew University, Jerusalem 91904, Israel. e-mail: shimons@cc.huji.ac.il

1. Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T. & Yamaguchi, A. *Nature* **443**, 173–179 (2006).
2. Seeger, M. A. *et al. Science* **313**, 1295–1298 (2006).
3. Dawson, R. J. P. & Locher, K. P. *Nature* **443**, 180–185 (2006).
4. Murakami, S., Nakashima, R., Yamashita, E. & Yamaguchi, A. *Nature* **419**, 587–593 (2002).
5. Yu, E. W., McDermott, G., Zgurskaya, H. I., Nikaido, H. & Koshland, D. E. Jr *Science* **300**, 976–980 (2003).
6. Davidson, A. L. & Chen, J. *Annu. Rev. Biochem.* **73**, 241–268 (2004).
7. Smith, P. C. *et al. Mol. Cell* **10**, 139–149 (2002).
8. Locher, K. P., Lee, A. T. & Rees, D. C. *Science* **296**, 1091–1098 (2002).
9. Chang, G. J. *Mol. Biol.* **330**, 419–430 (2003).
10. Reyes, C. L. & Chang, G. *Science* **308**, 1028–1031 (2005).

OBITUARY

James A. Van Allen (1914–2006)

Magnetospheric and space physicist, discoverer of the eponymous belts.

James A. Van Allen, who died in Iowa City on 9 August, was a native of Iowa — he was born in the small town of Mount Pleasant on 7 September 1914 — and lived and worked there for most of his life. But his influence, as a founding father of magnetospheric physics and world-renowned leader in the scientific exploration of space, extended far beyond the boundaries of his home state.

After attending both high school and the Iowa Wesleyan College in Mount Pleasant, Van Allen went to the University of Iowa in Iowa City, where he graduated with a PhD in physics in 1939. Following a spell in the department of terrestrial magnetism at the Carnegie Institution in Washington DC, Van Allen moved shortly after the outbreak of the Second World War to the Applied Physics Laboratory of Johns Hopkins University in Silver Springs, Maryland. There he worked to develop proximity fuses that allowed anti-aircraft shells to detonate automatically when reflected radio waves indicated the approach of a target. When the development phase of this project was finished, he was commissioned as a lieutenant in the US Navy and sent to the Pacific to test the fuses under combat conditions. For this assignment, he received four combat stars.

After the war, back initially at Johns Hopkins, and from 1951 as head of physics at the University of Iowa, Van Allen pioneered the use of Geiger tubes to measure cosmic rays at high altitudes. These tubes were initially attached to captured German V-2 and US Aerobee rockets, but Van Allen soon developed his own low-cost, balloon-borne rocket — which he called a rockoon — to carry out the studies at similar altitudes. The first comprehensive latitudinal survey of cosmic-ray intensities above Earth's atmosphere followed, using a series of rockets launched from ships as part of the 1957–58 International Geophysical Year.

During this period, Van Allen was also engaged in the development of an artificial satellite as part of the navy's Vanguard programme. On 4 October 1957, before these plans came to fruition, the Soviet Union launched the world's first satellite, Sputnik 1. In the ensuing frantic catch-up effort, Van Allen was asked to put his Geiger-tube instrumentation on a spacecraft, Explorer 1, to be carried by a rocket developed by Wernher von Braun. The spacecraft launched successfully on 31 January 1958.

Strapped to Explorer 1, the Geiger tube produced what seemed to be erratic results.

When passing over terrestrial receiving stations, it sometimes transmitted the expected cosmic-ray counting rates, and at other times did not count at all. A similar instrument on the Explorer 3 spacecraft carried a tape recorder that could record the counting rate for an entire orbit. This revealed that the periods of apparent non-operation were caused by extremely intense charged-particle radiation, thousands of times more intense than the cosmic-ray background, that were simply saturating the counter. The Van Allen belts — two doughnut-shaped regions of energetic charged particles surrounding Earth — had been discovered.

This entirely unexpected find attracted worldwide attention. Van Allen and his colleagues soon showed that the particles of the radiation belt — mainly protons in the inner belt and electrons in the outer belt — were trapped in stable, long-lived orbits by Earth's magnetic field. Over the nearly 50 years since the discovery of the belts, the study of this trapped radiation and its many effects, such as magnetic storms and auroral light emissions, has blossomed into a wide field known as magnetospheric physics.

Not content with his pioneering work on Earth's magnetosphere, Van Allen extended his research to other planets. He and his colleagues built similar Geiger-tube instrumentation for the Mariner 2 and 4 flights to Venus and Mars, respectively, and the Pioneer 10 and 11 flights to Jupiter and Saturn. Although Venus and Mars did not have radiation belts, Jupiter and Saturn were shown to have very large magnetospheres. As Pioneers 10 and 11 moved outward from the Sun, they also performed the first studies of the entry of cosmic rays into the heliosphere. This is the region around the Sun in which the solar wind — the stream of charged particles given off by its surface — has a controlling influence.

During his career, Van Allen published prolifically, producing more than 275 research papers and several books, one of which, *Origins of Magnetospheric Physics*, gives a detailed account of the early days of that field. He received many honours and prizes for his pioneering work in the field of



PICTORIAL PARADE/GETTY

space research, including the National Medal of Science in 1987, and the Crafoord prize — awarded by the Royal Swedish Academy of Sciences in disciplines outside the Nobel prize categories — in 1989.

Van Allen was also an outstanding teacher and mentor. He routinely taught a large, very popular undergraduate course in astronomy, and placed a strong emphasis on student involvement in his research. Many of his 34 PhD and 48 MSc students went on to become leaders in magnetospheric and space physics. That in turn gave university-based research a leading role in the US space programme, a role that endures today.

Van Allen's influence on his profession was immense. He served on many local and national scientific advisory committees. A strong proponent of robotic spacecraft, he became a persistent critic of human spaceflight, especially of the space shuttle and the International Space Station. He regarded these missions as inordinately expensive relative to the results obtained.

For his qualities of humanity and leadership in space science, James Van Allen will be missed by all who knew and worked with him.

Donald A. Gurnett

Donald A. Gurnett is in the Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242-1479, USA.
e-mail: donald-gurnett@uiowa.edu

BRIEF COMMUNICATIONS

Policing of oncogene activity by p53

Oncogenes, rather than DNA damage, may provide the key signal to p53 to trigger tumour suppression.

The tumour-suppressor protein p53 provides the most important genetic defence against cancer¹ and is activated in response to DNA damage and to oncogenic signalling, both of which occur almost universally in malignant tumours. But the relative contribution of these two pathways in inducing p53-dependent protection against cancer is unclear. Here we show that p53-dependent protection against cancer is lost in mice that have been genetically manipulated so that their p53 is activated in response to DNA damage but not to oncogenic signalling. We conclude that oncogenic signalling is the critical event that elicits p53-dependent protection and that the DNA-damage stimulus is less important.

DNA damage and oncogenic signalling are communicated to p53 through separate routes, which are, respectively, a p53-phosphorylation cascade that involves the ATM/Chk2/ATR/Chk1 series of kinases, and a p53-stabilization pathway that requires the tumour-suppressor protein ARF and the ubiquitin-ligase MDM2 (ref. 1).

To investigate the role of oncogenic signalling in p53-mediated protection against cancer, we used mice with two genetically engineered traits: one had no *ARF* allele (*ARF*^{null} mice)² and the other had a 'super' *p53* allele³ (*p53*^{super} mice; these mice carry a single additional transgenic copy of the intact *p53* gene, which behaves in the same way as endogenous *p53*). Compared with wild-type mice (*p53*^{wt} mice), which have just two copies of *p53*, the *p53*^{super} mice have additional protection against cancer development². This experimental system is therefore well suited for quantifying p53-dependent protection against cancer. (See supplementary information for methods.)

Before analysing their susceptibility to cancer, we confirmed that *ARF*^{null} mice respond normally to DNA damage^{4,5} by showing that apoptosis of their thymocytes after irradiation was unaffected (Fig. 1a). We found that mice with the *p53*^{super} allele showed the same enhancement of apoptosis irrespective of

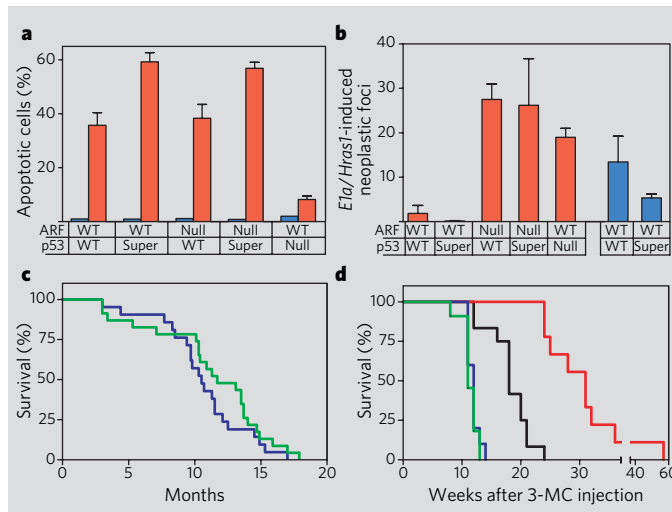


Figure 1 | ARF is necessary for tumour suppression by p53. **a**, The p53-dependent DNA-damage response *in vivo* does not depend on ARF. Mice ($n = 3$ per genotype) were irradiated (10 Gy) and the percentage of apoptotic thymocytes was determined 3 h later. Blue bars, non-irradiated controls; red bars, cells from irradiated mice. **b**, ARF is essential for the defensive response to oncogenic signals. Primary embryonic fibroblasts (from $n = 2$ embryos per genotype) were retrovirally transduced with *E1a* and oncogenic *Hras1* and plated as indicated, and the number of resulting neoplastic foci was scored. Red bars, 2,000 cells; blue bars, 50,000 cells. WT, wild type. **c**, Lifespans of *ARF*^{null}/*p53*^{wt} and *ARF*^{null}/*p53*^{super} mice ($n = 21$ (blue) and $n = 23$ (green) per genotype, respectively) were not significantly different (log rank test, $P = 0.26$). **d**, Mice of genotype *ARF*^{wt}/*p53*^{wt} ($n = 12$; black), *ARF*^{wt}/*p53*^{super} ($n = 9$; red), *ARF*^{null}/*p53*^{wt} ($n = 10$; blue) or *ARF*^{null}/*p53*^{super} ($n = 11$; green) were treated with the DNA-damaging agent 3-methyl cholanthrene (3-MC) and monitored for the development of fibrosarcomas. The protection against tumour development provided by the *p53*^{super} allele in the presence of ARF disappears in the absence of ARF; (see also supplementary information).

whether ARF was present or absent (Fig. 1a). However, *ARF*^{null} cells were unable to respond effectively to oncogenic signalling^{6–8} and underwent neoplastic transformation by oncogenes *in vitro*, irrespective of the presence or absence of the *p53*^{super} allele (Fig. 1b).

As p53 responds normally to DNA damage in the absence of ARF, we reasoned that *p53*^{super} might provide some protection against tumour development *in vivo*, even without the ability to detect oncogenic signalling. However, we found that *p53*^{super}/*ARF*^{null} mice succumbed to spontaneous tumours at the same rate as *p53*^{wt}/*ARF*^{null} mice (Fig. 1c), producing the same profile of sarcomas, lymphomas and histiocytic sarcomas (results not shown).

We also treated *p53*^{super}/*ARF*^{null} and *p53*^{wt}/*ARF*^{null} mice with the DNA-damaging agent 3-methyl cholanthrene. This agent produces

DNA adducts and results in fibrosarcomas carrying oncogenic mutations in *ras* genes (see supplementary information). This carcinogenic protocol is highly sensitive to the functionality of p53, as indicated by the greater resistance to the agent of *p53*^{super} mice compared with *p53*^{wt} mice³. As with the spontaneous tumours, the extra gene dose of *p53* became irrelevant in the absence of ARF (Fig. 1d).

Together, our results indicate that the cancer-protective activity of p53 is abolished in the absence of ARF. We conclude that oncogenic signalling is critical for triggering protection by p53, whereas activation of p53 as a result of DNA damage has a lesser impact on the ultimate development of tumours. Although there are differences in these pathways in mice and humans, our findings may also explain the high incidence of ARF loss in human cancers⁹, as well as the low incidence of mutations in the kinase enzymes of the p53-phosphorylation cascade¹⁰ that is induced by DNA damage.

Alejo Efeyan, Isabel Garcia-Cao, Daniel Herranz, Susana Velasco-Miguel, Manuel Serrano
Spanish National Cancer Centre (CNIO), Madrid 28029, Spain

e-mail: mserrano@cnio.es

- Harris, S. L. & Levine, A. J. *Oncogene* **24**, 2899–2908 (2005).
- Kamijo, T. et al. *Cell* **91**, 649–659 (1997).
- Garcia-Cao, I. et al. *EMBO J.* **21**, 6225–6235 (2002).
- Stott, F. J. et al. *EMBO J.* **17**, 5001–5014 (1998).
- Kamijo, T. et al. *Cancer Res.* **59**, 2464–2469 (1999).
- Palmero, I., Pantoja, C. & Serrano, M. *Nature* **395**, 125–126 (1998).
- Zindy, F. et al. *Genes Dev.* **12**, 2424–2433 (1998).
- de Stanchina, E. et al. *Genes Dev.* **12**, 2434–2442 (1998).
- Sharpless, N. E. *Mutat. Res.* **576**, 22–38 (2005).
- Bartek, J. & Lukas, J. *Cancer Cell* **3**, 421–429 (2003).

Supplementary information accompanies this communication on Nature's website.

Received 1 August; accepted 22 August 2006.

Competing financial interests: declared none.

doi:10.1038/443159a

BRIEF COMMUNICATIONS ARISING online
www.nature.com/bca see Nature contents.

ARCHAEOLOGY

Progress and pitfalls in radiocarbon dating

Arising from: P. Mellars *Nature* **439**, 931–935 (2006)

The review by Mellars¹ provides an archaeological perspective on some of the latest developments in radiocarbon dating of bone. However, his presentation, calibration and interpretation of radiocarbon ages are misleading in some cases, and important technical advances in the pretreatment of other sample types that have increased dating accuracy have been overlooked. These considerations may undermine the conclusions drawn about past human dispersals.

As Mellars points out¹, ultrafiltration of bone samples can yield final age determinations that are significantly different from ages derived by traditional methods^{1,2}. The latest results from ultrafiltration show that the ages originally reported from archaeological contexts should be considered as only minimum estimates. Although Mellars says that bones are the most widely available materials for dating¹, most (75%) of the ages plotted in Fig. 3 of ref. 1 are based on charcoal (P. Mellars, personal communication). As with bone, important developments have taken place in the preparation of charcoal for radiocarbon dating. Advances in the pretreatment and graphitization of charcoal³ have extended its dating range to at least 55,000 yr before present (BP) and make it probable that many

of the original age determinations for charcoal are minima. Unfortunately, Mellars reconstructs routes and rates of modern human dispersal across Eurasia from published ages of samples that had not been pretreated using the new methods.

Mellars also erroneously refers to NotCal04 (ref. 4) as a 'calibration'; however, NotCal04 is a summary of data sets that extend beyond the 26,000 calibrated years BP limit of the internationally accepted radiocarbon curve, IntCal04, and is not a calibration curve (hence the name⁴). The records that contribute to the NotCal04 comparison curve deviate by up to several millennia from one another^{5–8} (Fig. 1). Significant variability in atmospheric radiocarbon content is also indicated by a recently produced data set⁹. Some of this variability is due to uncertainties in the dating of individual records, but features common to all records will include, for example, short-lived excursions in the Earth's magnetic field⁸.

Atmospheric ¹⁴C variability has not followed a simple, smooth pattern, as suggested by Mellars¹. Instead, smoothing took place during the statistical analysis of these data sets to develop the NotCal04 mean best-fit line. By using the mid-point of the mean best-fit line, Mellars artificially improves the apparent precision of

calibrated ages in his Fig. 3; even 'infinitely' old ages are reported with improved precision, whereas calibration almost invariably results in age ranges that are significantly larger than the radiocarbon measurement error¹⁰.

We appreciate that Mellars' review was restricted to radiocarbon dating, principally of bone, but it is recommended practice that multiple methods and materials should be investigated to avoid any possible pitfalls that might be associated with a single technique or sample type. Several numerical-age methods are now available for the dating of bone¹¹, sediments and artefacts¹² that complement one another and add much-needed rigour to radiocarbon chronologies. We recommend that the accepted nomenclature for ages should be strictly followed to avoid confusion¹³: radiocarbon ages should be reported as 'BP', time before present (where 'present' is AD 1950); calibrated ages as 'cal. yr BP'; and ages from alternative dating methods in calendar years.

Inevitably, researchers will need to compare finds older than 26,000 cal. yr BP to other records measured on calendar timescales. When ages are 'calibrated', it must be specified which curve has been used. Meanwhile, radiocarbon ages should be reported in their uncorrected form for recalculation when an accepted calibration curve is developed.

Only by taking these considerations into account can we hope to develop robust models for human dispersal.

Chris S. M. Turney, Richard G. Roberts, Zenobia Jacobs

GeoQuEST Research Centre, School of Earth and Environmental Sciences, University of Wollongong, Wollongong, New South Wales 2522, Australia
e-mail: turney@uow.edu.au

1. Mellars, P. *Nature* **439**, 931–935 (2006).
2. Higham, T. F. G., Jacobi, R. M. & Ramsey, C. B. *Radiocarbon* **48**, 179–195 (2006).
3. Turney, C. S. M. *et al. Quat. Res.* **55**, 3–13 (2001).
4. van der Plicht, J. *et al. Radiocarbon* **46**, 1225–1238 (2004).
5. Hughen, K. *et al. Science* **303**, 202–207 (2004).
6. Kitagawa, H. & van der Plicht, J. *Radiocarbon* **42**, 369–380 (2000).
7. Beck, J. W. *et al. Science* **292**, 2453–2458 (2001).
8. Voelker, A. H. L., Grootes, P. M., Nadeau, M.-J. & Sarinthein, M. *Radiocarbon* **42**, 437–452 (2000).
9. Fairbanks, R. G. *et al. Quat. Sci. Rev.* **24**, 1781–1796 (2005).
10. Guilderson, T. P., Reimer, P. J. & Brown, T. A. *Science* **307**, 362–364 (2005).
11. Eggin, S. M. *et al. Quat. Sci. Rev.* **24**, 2523–2538 (2005).
12. Roberts, R. G. *Radiat. Meas.* **27**, 819–892 (1997).
13. Colman, S. M., Pierce, K. L. & Birkeland, P. W. *Quat. Res.* **28**, 314–319 (1987).

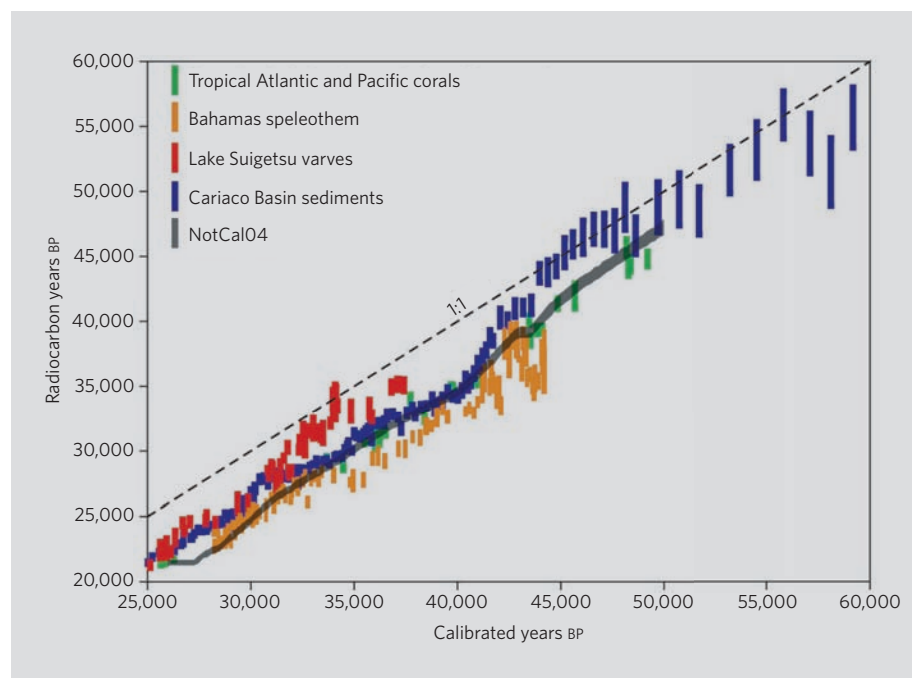


Figure 1 | Selected radiocarbon comparison curves for the period 25,000–60,000 calibrated years before present (BP). The graph shows the NotCal04 best-fit curve⁴, and Cariaco Basin sediments⁵, Lake Suigetsu varves⁶, the Bahamas speleothem⁷, and tropical Atlantic and Pacific corals⁹ radiocarbon data sets. Age ranges at 1 σ uncertainty (including those for the NotCal04 curve) are plotted only on the 'radiocarbon years BP' axis.

doi:10.1038/nature05214

ARCHAEOLOGY

Mellars replies

Replying to: C. S. M. Turney, R. G. Roberts & Z. Jacobs *Nature* **443**, doi:10.1038/nature05214 (2006)

Turney *et al.*¹ offer a useful summary of their own observations on new developments in the dating of charcoal samples. There are, however, several errors or misunderstandings in their reading of my review².

I was careful to describe the NotCal04 calibration study³ explicitly as a “best-estimation” curve (as the authors³ also describe it). But, even if there were an accepted calibration curve for the period before 26,000 cal. yr BP (calibrated years before present), there are still problems with any ‘authorized’ calibration curve because these can soon become out of date. As I pointed out in the last sentence of my review², the implications of such studies for the interpretation of the human archaeological and evolutionary record need to be kept under active and vigilant review.

Contrary to the assertion by Turney *et al.*¹, I did not suggest that atmospheric ¹⁴C variability has followed a simple, smooth pattern, but noted instead that the new calibration curves (from the Cariaco Basin, Atlantic and Pacific coral samples, the Socotra island speleothem, and so on) follow a “relatively” smooth pattern, “apparently” without any of the sudden and

aberrant oscillations in the atmospheric ¹⁴C content that had been claimed in some earlier studies. Important revisions^{4,5} have now been made to the latest calibration curves for both the Bahamas speleothem record⁶ and for the Cariaco Basin calibration curve⁷, which now show significantly smaller deviations from other calibration curves⁸ and render out of date the data plotted in Fig. 1 by Turney *et al.*¹. If all these revisions^{4,5} in the different calibration curves were incorporated into a new best-estimation curve, this curve would now be even tighter and would conform closely to the already published NotCal04 curve. I made a point of stressing² the potential impact of geomagnetic events, such as the Lake Mono and Laschamp excursions, for which the full impact on the terrestrial (as opposed to marine) radiocarbon records have still to be clearly elucidated.

I emphasized (Fig. 3 of ref. 2) that all radiocarbon dates (whether determined on bone or charcoal) should be regarded essentially as minimum figures, owing to the potential effects of contamination by recent, intrusive carbon. And I also made a point of presenting both the calibrated and uncalibrated versions

of the dates I used. (If I had plotted dates only for charcoal samples that had been processed by newly developed methods, such as the ABOX technique, there would have been almost no dates to plot.) I did not claim to improve the precision of ‘greater than’ age estimates, although I noted that calibration increases the minimum age of such estimates.

The implications of regarding all of these dates as essentially minimum age estimates must be kept clearly in mind in any applications of ¹⁴C dating to modern human dispersals, as I took pains to stress². Even so, the very marked geographical patterns visible in my Fig. 3 of the combined, best-available charcoal and bone dates should not be dismissed, even by the most pessimistic critics of the present state of radiocarbon dating in prehistory.

Paul Mellars

Department of Archaeology, Cambridge University, Cambridge CB2 3DZ, UK
e-mail: pam59@cam.ac.uk

1. Turney, C. S. M., Roberts, R. G. & Jacobs, Z. *Nature* **443**, doi:10.1038/nature05214 (2006).
2. Mellars, P. *Nature* **439**, 931–935 (2006).
3. van der Plicht, J. *et al. Radiocarbon* **46**, 1225–1238 (2004).
4. Beck, J. W. *Nineteenth Int. Radiocarbon Conf. April 3–7 abstr.* (Oxford, UK, 2006).
5. Hughen, K. *Nineteenth Int. Radiocarbon Conf. April 3–7 abstr.* (Oxford, UK, 2006).
6. Beck, J. W. *et al. Science* **292**, 2453–2458 (2001).
7. Hughen, K. *et al. Science* **303**, 202–207 (2004).
8. Fairbanks, R. G. *et al. Quat. Sci. Rev.* **24**, 1781–1796 (2005).

doi:10.1038/nature05215

Variations in solar luminosity and their effect on the Earth's climate

P. Foukal¹, C. Fröhlich², H. Spruit³ & T. M. L. Wigley⁴

Variations in the Sun's total energy output (luminosity) are caused by changing dark (sunspot) and bright structures on the solar disk during the 11-year sunspot cycle. The variations measured from spacecraft since 1978 are too small to have contributed appreciably to accelerated global warming over the past 30 years. In this Review, we show that detailed analysis of these small output variations has greatly advanced our understanding of solar luminosity change, and this new understanding indicates that brightening of the Sun is unlikely to have had a significant influence on global warming since the seventeenth century. Additional climate forcing by changes in the Sun's output of ultraviolet light, and of magnetized plasmas, cannot be ruled out. The suggested mechanisms are, however, too complex to evaluate meaningfully at present.

The possible effect on climate of variations in the Sun's total power output, or luminosity, has interested the scientific and financial communities since at least William Herschel's eighteenth-century study of correlations between sunspots and British grain prices. A proper measure for the Sun's output in this context is the wavelength-integrated radiation flux illuminating the Earth at its average distance from the Sun, called the total solar irradiance (TSI). Uncertainties imposed by the Earth's atmosphere delayed precise determination of variations in TSI until satellite-borne radiometry became available about a quarter of a century ago. Although the TSI fluctuations discovered so far through these measurements are too small to have a significant effect on climate¹, their analysis has now led to a robust understanding of solar luminosity variations, and new insights into the influence of magnetic fields on stellar convection. This advance in understanding is important, because it allows us to use the relatively short time series of space-borne measurements, along with other kinds of solar observations, to draw conclusions about the probable behaviour of TSI over past centuries and millennia, the timescales of more direct interest to climate studies.

From the theory of stellar evolution, we know that the Sun has gradually brightened by about 30% since it settled down to steady nuclear burning of hydrogen roughly 4.5 billion years ago. In spite of this reduced luminosity, the early Earth avoided completely freezing over, most probably because of much higher concentrations of greenhouse gases, such as methane or carbon dioxide (CO₂) (ref. 2). The more recent glacial/interglacial cycles provide us with an example of climate response over the shorter timescales of tens of thousands to millions—rather than billions—of years. The driving force here appears to be mainly variations in insolation, that is, the latitudinal and seasonal distribution of solar energy incident on the Earth, rather than in TSI itself. Such variations are thought to be caused by cyclic and predictable changes in the Earth's orbital parameters, the primary forcing effect being amplified by feedbacks involving CO₂ (ref. 3).

Here we review recent advances in our understanding of variations in TSI itself, which affect the planet's energy balance directly. This driving is relatively straightforward compared to possible climate

driving due to the variable solar output of ultraviolet light, and of magnetized plasmas in the solar wind. These more speculative possibilities are difficult to assess at present.

Sunspots and bright faculae cause TSI variations

The first convincing evidence of TSI variation was obtained by ACRIM (Active Cavity Radiometer Irradiance Monitor), an electrical substitution radiometer flown on the Solar Maximum Mission (SMM) satellite in 1980 (ref. 4). The data showed clearly how TSI decreases from day to day as a dark sunspot forms or rotates onto the solar disk, then recovers as the spot's projected area declines. Those first historic measurements from ACRIM, and also from the somewhat less precise Nimbus 7 radiometry beginning in late 1978, initiated a continuous record of TSI compiled from several spacecraft⁵.

This record (Fig. 1) shows a day-to-day variation of about 0.3% peak-to-peak amplitude most evident near maxima of the solar activity cycle. The high-frequency signal is superposed on an 11-year variation, of approximately 0.05–0.07% amplitude, which peaks around maxima of sunspot number. Brightening of the Sun with increasing sunspot number would be difficult to explain if dark magnetic sunspots were the only structures visible on the solar disk. In fact, as seen in Fig. 2, the Sun's photosphere (the atmospheric layer we see in broad-band visible light) is also flecked with many bright magnetic areas called faculae. They are immediately apparent where concentrated in bright clumps around sunspot groups. In a less conspicuous way, they are widespread also in the so-called magnetic network, which covers most of the solar surface (see, for example, ref. 6). We refer below to the bright clumps and the magnetic network collectively as the bright component of the surface magnetic field, or as 'magnetic brightenings'.

Each such dark or bright structure contributes a TSI variation equal to the product of its projected area and its photometric contrast relative to the adjacent, undisturbed photospheric disk. Empirical models reconstruct a TSI variation by adding up these contributions using records of the changing areas of the spots and magnetic brightenings. These models can account for over 80% of the variance in the irradiance time series^{7–11} (Fig. 3). This shows that additional

¹Heliophysics, Inc., Nahant, Massachusetts 01908, USA. ²Physikalisch-Meteorologisches Observatorium Davos, World Radiation Center, CH-7260 Davos Dorf, Switzerland.

³Max Planck Institute for Astrophysics, D-85741 Garching, Germany. ⁴National Center for Atmospheric Research, Boulder, Colorado 80307-3000, USA.

contributions to the TSI variation from photospheric temperature variations outside spots and magnetic brightenings must be small. Apparently, neither the darkness of spots nor the increased flux from the bright component is cancelled by compensating variations in solar brightness elsewhere on the disk. This has historically caused significant debate in the literature, because the photometric accuracy of solar images is insufficient to detect such cancelling contributions if they are spread over large areas. Their possible presence is thus hard to disprove by more direct means.

Physical mechanisms of luminosity variation

The observations show that magnetic structures reduce or increase the local radiation flux for as long as they exist at the surface. They present (in the case of sunspots, for example) a local magnetic obstruction to the flow of heat through the Sun. One may ask why such a disturbance in the heat flow does not simply readjust ('shunt') around the obstacle. If this were the case, it would not cause a TSI variation at all. Several mechanisms that might store (for spots) or release (for the bright component) the energy involved in irradiance variations have been suggested in response to this conceptual problem. For instance, the 'missing' radiative flux of a spot might, in principle, be stored in intensifying its strong magnetic field¹². But this storage would only occur while the spot's magnetic energy is increasing, whereas the amplitude of sunspot-induced dips in TSI depends only on changes of the spot's projected area, not on whether its true area is growing or decaying, or on the magnetic field strength¹³. More broadly, this observational constraint seems to disqualify any mechanisms that depend on plasma flows¹⁴ or other phenomena associated with growth of a spot. Magnetic 'stirring', a proposed enhanced entrainment of heat associated with rising magnetic flux tubes¹⁵, seems to play little part, given the close amplitude match between measured and reconstructed TSI variations.

Modelling of sunspots using current magnetohydrodynamic codes¹⁶ confirms that their darkness can be understood as thermal 'plugs' in which intense vertical magnetic fields divert heat flow from deeper layers^{17,18}. Magnetic brightenings act in the opposite sense, as local thermal 'leaks'^{19,20}. They are bright because the effect of their smaller-scale magnetic field is to form small depressions in the photospheric surface, enabling radiation from lower and hotter

atmospheric layers to escape more easily^{19,21}. These depressions have now been imaged directly in the highest-resolution pictures of the photosphere²² (Fig. 2), and are reproduced in detail by realistic three-dimensional radiative magnetohydrodynamic simulations of these small magnetic structures²³.

Calculations with such a thermal impedance model of solar luminosity variation (Fig. 4) show that the heat flux diverted by a spot is not merely shunted aside to reappear nearby. Instead, the thermal perturbation spreads horizontally and back down into the Sun, much more quickly than it can be radiated from the photosphere. As a consequence, the 'blocked heat flux' effectively remains stored in the Sun for hundreds of millennia instead of appearing elsewhere on the surface, in agreement with the observation that 'bright rings' around spots are weak or absent^{24,25}. This rapid redistribution is also in agreement with the absence of 'thermal shadows' preceding spot emergence²⁶, and the absence of phase differences between the irradiance dip and the appearance of a spot's magnetic field at the photosphere¹⁸. The thermal plug model predicts that the Sun's luminosity immediately drops by a factor equal to the product of the spot's cross-sectional area and its brightness deficit, as observed. This factor is independent of the spot's growth or decay rate, also in agreement with observations¹³. The heat missing from the spot is stored as slightly increased internal and potential energy of the entire solar convective envelope. The predicted changes in solar radius accompanying changes in spot area are unobservably small²⁷. This is consistent with the increasingly tight upper limits on measured radius variations discussed below.

The Sun's remarkable ability to store, rather than quickly re-radiate, fluctuations in heat flow to the photosphere can be understood in terms of a simple illustration. Consider a satellite in temperature equilibrium with its surroundings in space. Then imagine that current flow through a resistor on the satellite is briefly turned on, sending a heat pulse spreading rapidly through the satellite at a rate determined by the thermal diffusivity of aluminium, its main constituent. New isotherms describing the somewhat warmer temperature distribution within the satellite will be relatively quickly established over this diffusive timescale. But the satellite is

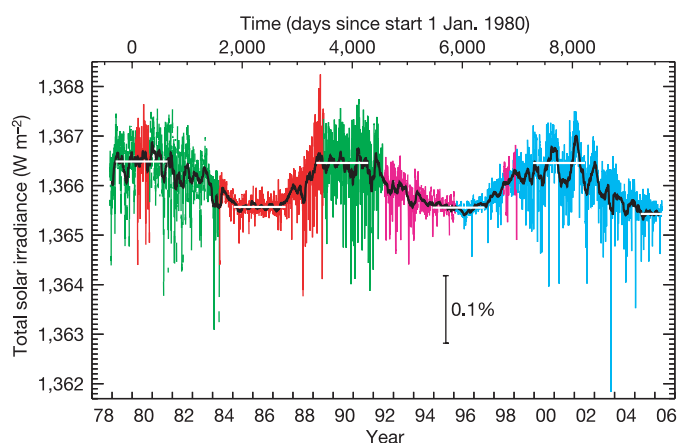


Figure 1 | Temporal variation of total solar irradiance, as measured by radiometers on several spacecraft since 1978. Daily measurements are shown with a 81-day running average. The TSI increases around the maxima of sunspot number that occurred near 1980, 1990 and 2001. The high-frequency variation is caused by the changing projected areas of spots and faculae on the solar disk as the Sun rotates on its axis in approximately 27 days. The TSI variation amplitudes of the three sunspot cycles shown (horizontal lines) are 0.92, 0.89 and 0.90 W m^{-2} , respectively, with an average minimum of $1,365.52 \pm 0.009 \text{ W m}^{-2}$ and differences of the minima from this average of $+0.051$, $+0.037$ and -0.089 W m^{-2} .

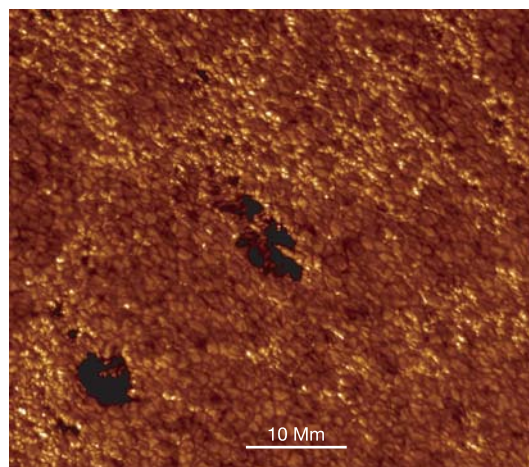


Figure 2 | An image of the solar photosphere (the surface of the solar disk seen in visible light), showing the structures responsible for the TSI variations. The granules, covering most of the area, are the convective flows carrying energy from the interior. They contribute the steady component of the irradiance received by the Earth. Magnetic structures are dark (sunspots) or bright (the small bright points called faculae) and contribute a component that varies with the sunspot cycle. Faculae show up especially towards the limb of the solar disk. Their contribution dominates over the dark spots, so that the Sun is slightly brighter at sunspot maximum. Scale bar, ten megametres. (Image taken with the Swedish 1-m Solar Telescope by B. De Pontieu on 16 June 2003.)

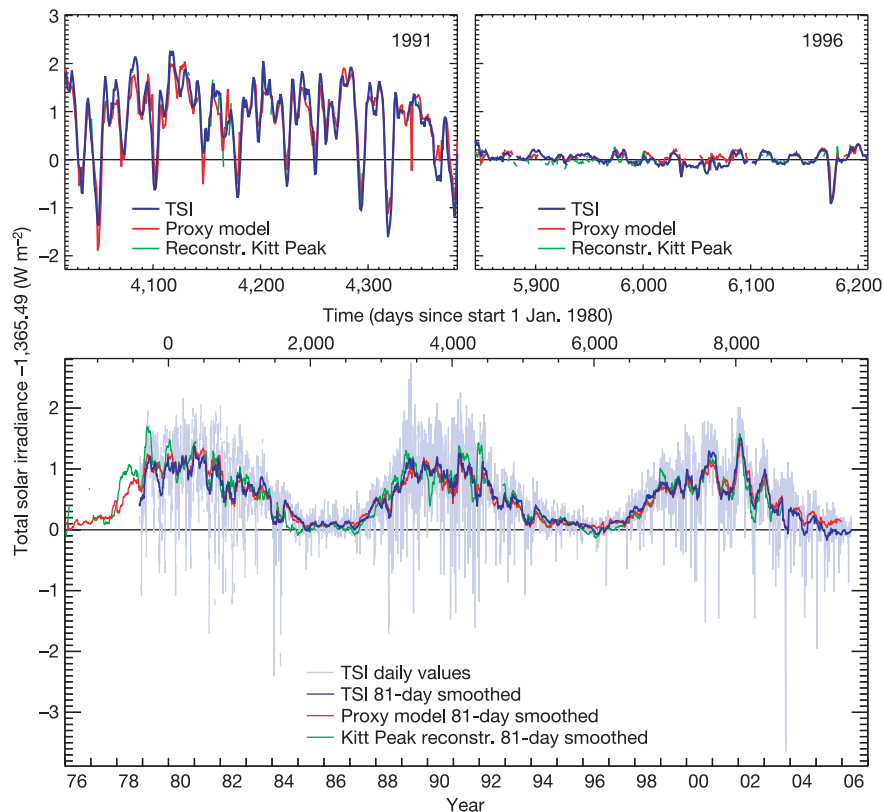


Figure 3 | Variations in total solar irradiance. Lower panel, a comparison of radiometrically measured TSI variation⁵ as daily values (light blue) and as 81-day smoothed curve (dark blue). Also plotted are TSI reconstructed from contributions of sunspots, faculae and network¹⁰ (red curve) and from Kitt Peak magnetograms¹¹. Differences between the red and green curves illustrate remaining uncertainties in the reconstruction of TSI variation over

the 11-year cycle. The upper panels illustrate fluctuations due to spots and magnetic brightenings moving across the solar disk on the 27-day solar rotational timescale, with daily TSI values on an expanded timescale during a solar maximum (1991) and minimum (1996). Again, the reconstructions are also plotted for comparison.

now warmer than before, and must eventually return to equilibrium with its space surroundings by radiating the heat liberated by the resistor. For a satellite, this timescale for radiative equilibration is about two orders of magnitude longer than the diffusive timescale, and until it is achieved, the resistor's heat pulse is stored in the satellite.

The solar equivalent of the thermal diffusion in a metal is thermal exchange by convective flows. The radiative equilibration timescale of the Sun's convection zone, about 100,000 years, is about six orders of magnitude longer than this exchange timescale, and storage of thermal disturbances is much more efficient than in the metal analogy. As a consequence, neither the excess emission in the bright

magnetic component nor the darkness of spots is expected to be compensated by measurable changes in photospheric temperature elsewhere on the disk.

These findings were originally obtained from calculations in which convection was approximated by turbulent diffusion. Recent three-dimensional hydrodynamical simulations, which accurately reproduce observation of solar convection near the photosphere²⁸, confirm the picture sketched above, of the Sun's convective envelope as a 'thermal superconductor' with a huge thermal inertia²⁷. We note that this thermal inertia explains not only why 'superficial' photospheric magnetic structures can cause net irradiance fluctuations, but also explains why additional variations with 'deeper' origins in the Sun have not been observed so far.

Variation in irradiance by other mechanisms

A number of observational techniques have been used to search for other, possibly quite different, modes of solar luminosity variation. For example, variations in global heat flux carried by large convective cells are predicted by certain idealized models. Neither these cells, nor other evidence for global temperature variations, have been found in precise photometric mapping of the photosphere^{29,30}. Gradients of photospheric heat flux in heliographic latitude have been reported, but further analysis indicates that these arise from inhomogeneities in the distribution of the network³¹, rather than from deeper-lying convective patterns, as originally believed. A slow increase of photospheric effective temperature outside magnetic structures has been reported from measurements of certain temperature-sensitive photospheric spectral lines since 1978, but the amplitude is very small—roughly equivalent to a TSI variation of 0.01% per decade (ref. 32).

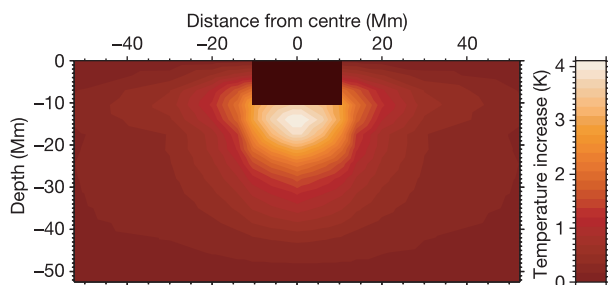


Figure 4 | Thermal disturbance caused by a sunspot, computed as a function of distance from it and depth into the Sun¹⁸. Brightness represents temperature difference compared with an unspotted model of the Sun. At the surface the increase almost vanishes: there is no 'bright ring' around the spot. Instead of reappearing at the surface, the thermal energy blocked by the dark spot remains stored in the solar interior^{17,18}.

Measurements of the solar diameter originally attracted attention because they seemed to offer a unique diagnostic of otherwise unobservable internal temperature variations. However, the amplitude of reported diameter variations has decreased with improved measurement techniques. Separate analyses of the data from the Michelson Doppler Imager (MDI) instrument on the SOHO spacecraft, using both heliometric and helioseismic techniques, show no evidence of secular trends in diameter, or variations attributable to the 11-year cycle, at the level of a few millarcseconds (a few kilometers on the Sun^{33,34}). This is consistent with, but ten times better than, upper limits from the best ground-based measurements of the Sun's diameter³⁵. Conflicting detections nearly two orders of magnitude larger continue to be cited, however^{36–38}. If diameter variations were detected within the tight upper limits they would be quite small relative to variations in TSI, indicating an origin in layers close below the surface^{27,39}. Thus, it is unlikely that diameter measurements can reveal deeper-lying sources of solar irradiance variation, as was originally hoped. In fact, they put strict limits on the presence of such sources. Their small amplitude also implies that changes in the size of the solar disk contribute negligibly to the TSI.

Part of the increased irradiance measured around sunspot maxima has been interpreted in some studies as evidence for enhanced temperatures extending deep inside the convective envelope⁴⁰. The resulting expansion of the envelope, however, would cause the frequencies of global sound wave oscillations to decrease⁴¹, whereas helioseismic observations show an increase at activity maximum⁴². Moreover, the frequency and wavenumber dependence of this effect show that it is not due to a change in the solar radius. Most of it is explained by the increased 'stiffness' of the solar atmosphere due to magnetic fields⁴³. A remaining small cooling seems consistent with the enhanced radiative losses from the surface due to the bright component⁴⁴, rather than with heating from below.

Overall, no significant evidence can be claimed at present for modes of irradiance variation other than the surface spottiness caused by magnetic fields. Studies of other stars support this

conclusion, as the reported photometric variability of Sun-like stars always correlates with magnetic activity level⁴⁵. So far, no cases have been reported of significant luminosity variations caused, for instance, by the random variations that one might expect from fluctuations of stellar convective efficiency.

This absence of other solar luminosity variation modes may seem remarkable. For example, the power required to drive the Sun's 11-year magnetic oscillation must ultimately be drawn from solar heat flow. Attempts to model such 'structural' luminosity variations³⁷ confirm the earlier finding that 11-year output variations can be produced provided that a perturbation is introduced in relatively shallow layers (a few per cent of the solar radius), where the thermal timescale is short^{27,39}. There is, however, no evidence at the observed surface itself of the kinds of magnetic fields that are assumed to exist just below it in these models. In addition, most dynamo models place the field generation deep inside or even below the convection zone, where thermal perturbations are severely attenuated because of the 100,000-year radiative relaxation time of these more massive layers. Such structural luminosity variations are thus unlikely to play a role on decadal or even centennial timescales.

Multi-decadal to millennial variations of TSI

If only photospheric magnetic structures influence TSI, what range of TSI variation might be expected from the behaviour of solar activity over recent centuries? One influential study⁴⁶ extrapolated this relationship between activity and TSI to activity levels lower than any directly observed on the Sun—levels where any vestige of even the weakest photospheric magnetic structures is removed. At the time, it was believed that such ultra-quiet behaviour had been observed in certain Sun-like stars. This extrapolation introduced an additional, low-frequency TSI variation having an amplitude about three times larger than the present 11-year variation expected from spots and the bright magnetic component.

The apparent success of this model in driving climate variations that appeared to match observed variations attracted wide attention. However, this study has since been retracted⁴⁷, partly because the stars it was based on have since been shown to be less Sun-like than was originally believed⁴⁸, and also because it implied that the magnetic network must have been absent before the beginning of the twentieth century. This disagrees with archival solar images⁴⁹. Observables invoked in other reconstructions—for example, sunspot cycle length^{50,51}, the smoothed activity level^{46,51}, or geomagnetic indices⁵²—lack a demonstrated connection to TSI variation, and such models remain speculative.

The possibility of larger irradiance variations generated by higher past solar activity levels has received less attention. In the present Sun, the photometric effects of the bright and dark magnetic components partially cancel, so higher levels of irradiance variation could result if this balance were to change⁵³. In fact, the ratio of sunspot to facular area is observed to increase at the highest solar activity, such as during cycle 19 peaking around 1957. Such spot dominance also seems required to explain the photometric variability on younger and more active stars that are otherwise similar to the Sun⁴⁵. Recent reconstructions from ¹⁴C records disagree on whether the Sun might have experienced extended episodes of very high activity in the past 10,000 years, much above that encountered in the recent past^{54,55}. The intriguing possibility of larger luminosity variations by such a 'hyperactive' Sun deserves closer attention.

The climate record and possible TSI variations

Additional evidence on past TSI variations might, in principle, be gleaned from reconstructions of the Earth's climate record. But the relationship between TSI variations and climate is complex, being significantly affected by uncertainties in the climate sensitivity, the large thermal inertia of the oceans, and the effects of non-solar forcings such as volcanism and human-induced factors.

Figure 5a shows seven reconstructions of Northern Hemisphere

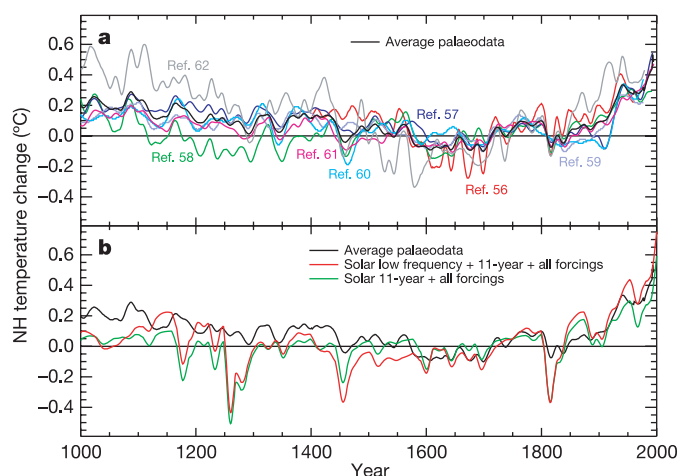


Figure 5 | Northern Hemisphere temperatures compared to model results. **a**, Variations of Northern Hemisphere (NH) mean temperature over the past millennium as reconstructed in seven recent studies^{56–62}. The plots were constructed from the original data, low-pass-filtered with a 30-year, 19-term gaussian filter. The black line shows the unweighted mean of the individual smoothed curves. All data are zeroed over 1601–1900. **b**, Comparison of the reconstructed Northern Hemisphere temperature for the past millennium with the modelled temperature changes including volcanic⁶⁸ and anthropogenic forcing⁶⁴, with TSI forcing by solar cycle amplitude (green curve) and with additional low-frequency variations (red curve); TSI data are from ref. 69, extended using ¹⁰Be (ref. 70). The climate sensitivity assumed is a central value of 2.6 °C equilibrium warming for a CO₂ doubling. The filtering and zeroing of the model results is the same as used for the temperature reconstructions.

temperature for the past millennium^{56–62}, smoothed with a 30-year low-pass filter, and their unweighted mean. The mean curve shows a gradual cooling to around 1700, followed by a warming that has accelerated in the twentieth century, resulting in a shape widely described as a ‘hockey stick’. Figure 5b compares the average temperature curve with climate model simulations of Northern Hemisphere temperature. We use the upwelling-diffusion energy balance climate model used by IPCC^{63–67} for these model simulations. In the present context, models of this type have advantages over more complex atmosphere–ocean general circulation models (see, for example, ref. 67), not least because the climate sensitivity can be specified by the user.

The present simulations consider TSI forcing combined with volcanic⁶⁸ and anthropogenic contributions^{63,64}. We consider that TSI variations are produced only by the net effect of dark and bright magnetic contributions (from ref. 69). As explained in previous sections, this is the only physically defensible TSI forcing at present. The TSI data were extended to 1000 AD using the ¹⁰Be isotope record⁷⁰.

It can be seen from Fig. 5b that the observed cooling towards a minimum around the seventeenth century is hardly significant in the modelled temperatures using this small but realistic TSI variation (green curve). We also considered the case where the forcing by the dark and bright magnetic components was modified by adding a low-frequency component⁶⁹, amplifying the solar forcing range by 3 to 4 over this time period. In this case (red curve), the model cooling trend is closer to the reconstructed palaeotemperature trend. Further simulations with solar forcing alone show that the solar contribution to warming over the past 30 years is negligible⁴⁸. These results indicate that a TSI variation at least three times larger than is justified from present evidence is required to produce a climate variation of the amplitude suggested by the seventeenth-century cooling. A more detailed statistical comparison between the modelled and reconstructed curves is problematic because of uncertainties in the palaeotemperature reconstruction, in the TSI proxies used, and in the modelling of, for example, the volcanic contribution.

The use of the sunspot number and of radioisotopes as proxies of TSI variation deserves some comment. As pointed out above, the only defensible TSI variation at present arises from the net contribution of bright and dark photospheric magnetic structures. The areas of such structures are available from measurements dating back only to the beginning of the twentieth century. For the earlier period dating back to the beginning of telescope observations of the solar disk in the early seventeenth century, the rough level of solar activity is commonly measured using a somewhat arbitrary statistic based on the daily number of sunspots visible through a small telescope. This sunspot number, *R*, gives a fair estimate of TSI decrease due to spots. On average, it also estimates the TSI increase due to the bright component, as this tends to vary in rough proportion to the sunspot dimming. However, the relative contributions to TSI of the dark and bright components vary significantly between sunspot cycles, so the estimate of TSI obtained from use of *R* in any given cycle can be in error by as much as 30%. Nevertheless, the behaviour of *R* does provide very interesting information on the variation of TSI over centennial timescales back to the early seventeenth century.

Use of a radio-isotope as a proxy for TSI variation introduces additional uncertainties. The radio-isotopes are not produced in the Earth’s atmosphere directly by the Sun’s flux of energetic particles. For recent levels of solar activity, they are produced mainly by high-energy galactic cosmic rays, from outside the Solar System. Their modulation with the solar cycle is due to changes in shielding of these cosmic rays by the heliosphere (the solar wind). The efficiency of this shielding depends on solar plasma outflows from open magnetic fields in quiet regions and individual events such as flares and coronal mass ejections⁷¹. Although this shielding increases roughly with the general level of solar activity, it is only very loosely proportional to the areas of the dark and bright magnetic structures that drive TSI. In

view of this, it is unrealistic to expect a fixed relation between variations in ¹⁰Be production rate and TSI over the past millennium. The relation is complicated further by possible climate influences on the ¹⁰Be deposition rate⁷², causing errors in the inferred ¹⁰Be formation rate.

Evidence for the influence of the Sun on climate has recently been extended by a millennial-scale correlation found between North Atlantic climate (inferred from the properties of deep sea sediments) and solar activity (inferred from ¹⁰Be and ¹⁴C), extending back about 10,000 years (ref. 73). It remains to be seen whether the correlation is confined to the North Atlantic, or is global⁷⁴. No specific mechanism has been identified so far to generate millennial-scale solar irradiance variations.

Overall, we can find no evidence for solar luminosity variations of sufficient amplitude to drive significant climate variations on centennial, millennial and even million-year timescales. Better reconstructions of global temperature and solar activity will be required to investigate further the apparent relationships between climate and solar activity seen over the past millennium and through the Holocene, particularly if the signature of any solar influence is spatially restricted⁷⁵.

Less direct Sun–climate couplings driven by the Sun’s well-known variability in ultraviolet flux and in outputs of magnetized plasma might yet account for Sun–climate correlations that defy explanation by the direct influence of TSI variation considered here. The proposed indirect mechanisms^{76,77} are, however, complex, and involve subtle interactions between the troposphere, stratosphere and even higher layers of the Earth’s atmosphere that are much less well understood than the direct radiative forcing effect. Modelling of such interactions is proceeding rapidly, but incisive tests of the models will be required to achieve certainty. For instance, a recent reconstruction⁷⁸ shows that the time series of solar ultraviolet flux since 1915 differs substantially from the behaviour of TSI and exhibits a surprisingly poor correlation with global temperature. Clearly, more needs to be done on such reconstructions to help guide future modelling.

1. Wigley, T. M. L. & Raper, S. C. B. Climatic change due to solar irradiance changes. *Geophys. Res. Lett.* **17**, 2169–2172 (1990).
2. Kasting, J. F. Planetary atmospheres: Warming early Earth and Mars. *Science* **276**, 1213–1215 (1997).
3. Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in the Earth’s orbit: Pacemaker of the Ice Ages. *Science* **194**, 1121–1132 (1976).
4. Willson, R. C., Gulkis, S., Janssen, M., Hudson, H. S. & Chapman, G. A. Observations of solar irradiance variability. *Science* **211**, 700–702 (1981).
5. Fröhlich, C. Solar irradiance variability since 1978: Revision of the PMOD composite during solar cycle 21. *Space Sci. Rev.* (in the press); preprint at (http://ftp.pmodwrc.ch/pub/Claus/ISSI_WS2005/ISSI2005a_CF.pdf) (2006).
6. Foukal, P. *Solar Astrophysics* 2nd edn (Wiley-VCH, Weinheim, 2004).
7. Foukal, P., Bernasconi, P. & Walton, S. A model of solar irradiance variation using broad band measurements of facular photometric contrast. *Astrophys. J.* (in the press); preprint at (http://sd-www.jhuapl.edu/SBI/Pubs/Foukal_et_al_ApJ_Letters_2006.pdf) (2006).
8. Foukal, P. V. & Lean, J. Magnetic modulation of solar luminosity by photospheric activity. *Astrophys. J.* **328**, 347–357 (1988).
9. Chapman, G. A., Cookson, A. M. & Dobias, J. J. Variations in total solar irradiance during solar cycle 22. *J. Geophys. Res.* **101**, 13541–13548 (1996).
10. Fröhlich, C. & Lean, J. Solar radiative output and its variability: Evidence and mechanisms. *Astron. Astrophys. Rev.* **12**, 273–320 (2004).
11. Wenzler, T. *Reconstruction of Solar Irradiance Variations in Cycles 21–23 based on Surface Magnetic Fields*. PhD thesis, Eidgenössische Technische Hochschule, Zürich (2005).
12. Wilson, P. R. in *Physics of Sunspots* (eds Cram, L. & Thomas, J.) 83–97 (Sacramento Peak Observatory, Sunspot, New Mexico, 1982).
13. Zahid, H. J., Hudson, H. S. & Fröhlich, C. Total solar irradiance variation during rapid sunspot growth. *Sol. Phys.* **222**, 1–15 (2004).
14. Schatten, K. H., Mayr, H. G., Omidvar, K. & Maier, E. A hillock and cloud model for faculae. *Astrophys. J.* **311**, 460–473 (1986).
15. Parker, E. N. Theoretical properties of omega-loops in the convective zone of the sun. 2: The origin of enhanced solar irradiance. *Astrophys. J.* **440**, 415–420 (1995).
16. Schüssler, M. & Vögler, A. Magnetoconvection in a sunspot umbra. *Astrophys. J.* **641**, L73–L76 (2006).
17. Spruit, H. C. The flow of heat near a starspot. *Astron. Astrophys.* **108**, 356–360 (1982).

18. Foukal, P., Fowler, L. A. & Livshits, M. A thermal model of sunspot influence on solar luminosity. *Astrophys. J.* **267**, 863–871 (1983).
19. Spruit, H. C. Heat flow near obstacles in the solar convection zone. *Sol. Phys.* **55**, 3–34 (1977).
20. Chiang, W. H. & Foukal, P. V. The influence of faculae on sunspot heat blocking. *Sol. Phys.* **97**, 9–20 (1985).
21. Spruit, H. C. Pressure equilibrium and energy balance of small photospheric fluxtubes. *Sol. Phys.* **50**, 269–295 (1976).
22. Lites, B. W., Scharmer, G. B., Berger, T. E. & Title, A. M. Three-dimensional structure of the active region photosphere as revealed by high angular resolution. *Sol. Phys.* **221**, 65–84 (2004).
23. Carlsson, M., Stein, R. F., Nordlund, Å. & Scharmer, G. B. Observational manifestations of solar magnetoconvection: Center-to-limb variation. *Astrophys. J.* **610**, L137–L140, doi:10.1086/423305 (2004).
24. Fowler, L. A., Foukal, P. V. & Duvall, T. L. Jr Sunspot bright rings and the thermal diffusivity of solar convection. *Sol. Phys.* **84**, 33–44 (1983).
25. Rast, M. P. *et al.* Bright rings around sunspots. *Nature* **401**, 678–680 (1999).
26. Foukal, P. V. & Fowler, L. A. A photometric study of heat flow at the solar photosphere. *Astrophys. J.* **281**, 442–454 (1984).
27. Spruit, H. in *Sun in Time* (eds Sonnet, C., Giampapa, M. S. & Matthews, M. S.) 118–158 (Univ. Arizona Press, Tucson, 1991).
28. Rosenthal, C. S., Christensen-Dalsgaard, J., Nordlund, Å., Stein, R. F. & Trampedach, R. Convective contributions to the frequencies of solar oscillations. *Astron. Astrophys.* **351**, 689–700 (1999).
29. Petro, L. D., Foukal, P. V. & Kurucz, R. L. Photospheric limb-darkening signatures of global structure variations. *Sol. Phys.* **98**, 23–38 (1985).
30. Chiang, W.-H., Petro, L. D. & Foukal, P. V. A photometric search for solar giant convection cells. *Sol. Phys.* **110**, 129–138 (1987).
31. Woodard, M. F. & Libbrecht, K. G. Spatial and temporal variations in the solar brightness. *Sol. Phys.* **212**, 51–64 (2003).
32. Livingston, W. & Wallace, L. The Sun's immutable basal quiet atmosphere. *Sol. Phys.* **212**, 227–237 (2003).
33. Antia, H. M. Does the Sun shrink with increasing magnetic activity? *Astrophys. J.* **590**, 567–572 (2003).
34. Kuhn, J. R., Bush, R. I., Emilio, M. & Scherrer, P. H. On the constancy of the solar diameter II. *Astrophys. J.* **613**, 1241–1252 (2004).
35. Brown, T. M. & Christensen-Dalsgaard, J. Accurate determination of the solar photospheric radius. *Astrophys. J.* **500**, L195–L199 (1998).
36. Sofia, S. Global variability of the Sun. *Mem. Soc. Astron. Ital.* **69**, 531–537 (1998).
37. Sofia, S. Variations of total solar irradiance produced by structural changes of the solar interior. *Eos* **85**, 217–221 (2004).
38. Egidi, A. *et al.* High-precision measurements of the solar diameter and oblateness by the Solar Disk Sextant (SDS) Experiment. *Sol. Phys.* **235**, 407–418 (2006).
39. Gilliland, R. L. Theoretical interpretation of the variability of global solar properties. *Adv. Space Res.* **8**, 151–155 (1988).
40. Kuhn, J. R., Bush, R. I., Scherrer, P. & Scheick, X. The Sun's shape and brightness. *Nature* **392**, 155–157 (1998).
41. Goldreich, P., Murray, N., Willette, G. & Kumar, P. Implications of solar p-mode frequency shifts. *Astrophys. J.* **370**, 752–762 (1991).
42. Libbrecht, K. C. & Woodard, M. F. Solar-cycle effects on solar oscillation frequencies. *Nature* **345**, 779–782 (1990).
43. Dziembowski, W. A. & Goode, P. R. Sources of oscillation frequency increase with rising solar activity. *Astrophys. J.* **625**, 548–555 (2005).
44. Spruit, H. C. Origin of the torsional oscillation pattern of solar rotation. *Sol. Phys.* **213**, 1–21 (2003).
45. Radick, R. R., Lockwood, G. W., Skiff, B. A. & Baliunas, S. L. Patterns of variation among Sun-like stars. *Astrophys. J. Suppl.* **118**, 239–258 (1998).
46. Lean, J., Beer, J. & Bradley, R. Reconstruction of solar irradiance since 1610: Implications for climate change. *Geophys. Res. Lett.* **22**, 3195–3198 (1995).
47. Lean, J. L., Wang, Y.-M. & Sheeley, N. R. The effect of increasing solar activity on the Sun's total and open magnetic flux during multiple cycles: Implications for solar forcing of climate. *Geophys. Res. Lett.* **29**, 2224, doi:10.1029/2002GL015880 (2002).
48. Foukal, P., North, G. & Wigley, T. M. L. A stellar view on solar variations and climate. *Science* **306**, 68–69 (2004).
49. Foukal, P. & Milano, L. A measurement of the quiet network contribution to solar irradiance variation. *Geophys. Res. Lett.* **28**, 883–886 (2001).
50. Hoyt, D. V. & Schatten, K. H. A discussion of plausible solar irradiance variations, 1700–1992. *J. Geophys. Res.* **98**, 18895–18906 (1993).
51. Solanki, S. & Fligge, M. Solar irradiance since 1874 revisited. *Geophys. Res. Lett.* **25**, 341–344 (1998).
52. Lockwood, M., Stamper, R. & Wild, M. A doubling of the Sun's coronal magnetic field during the past 100 years. *Nature* **399**, 437–439 (1999).
53. Foukal, P. What determines the relative areas of spots and faculae on Sun-like stars? *Astron. Astrophys.* **500**, 958–965 (1998).
54. Solanki, S. K., Usoskin, I. G., Kromer, B., Schüssler, M. & Beer, J. Unusual activity of the Sun during recent decades compared to the previous 11,000 years. *Nature* **431**, 1084–1087 (2004).
55. Muscheler, R., Joos, F., Müller, S. & Snowball, I. Climate: How unusual is today's solar activity? *Nature* **436**, E4–E5 (2005).
56. Briffa, K. R. *et al.* Low-frequency temperature variations from a northern tree ring density network. *J. Geophys. Res.* **106**, 2929–2942, doi:10.1029/2000JD900617 (2001).
57. Crowley, T. J. Causes of climate change over the past 1000 years. *Science* **289**, 270–277 (2000).
58. Esper, J., Cook, E. R. & Schweingruber, F. H. Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. *Science* **295**, 2250–2253 (2002).
59. Jones, P. D., Osborn, T. J. & Briffa, K. R. The evolution of climate over the last millennium. *Science* **292**, 662–667 (2001).
60. Mann, M. E., Bradley, R. S. & Hughes, M. K. Northern Hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophys. Res. Lett.* **26**, 759–762 (1999).
61. Mann, M. E. & Jones, P. D. Global surface temperatures over the past two millennia. *Geophys. Res. Lett.* **30**, 1820, doi:10.1029/2003GL017814 (2003).
62. Moberg, A., Sonechkin, D., Holmgren, K., Datsenko, N. & Karlén, W. Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature* **433**, 613–617 (2005).
63. Wigley, T. M. L. & Raper, S. C. B. Interpretation of high projections for global-mean warming. *Science* **293**, 451–454 (2001).
64. Cubasch, U. & Meehl, G. in *Climate Change 2001: The Scientific Basis* (eds Houghton, J. T. *et al.*) 525–582 (Cambridge Univ. Press, Cambridge, UK, 2001).
65. Raper, S. C. B., Gregory, J. M. & Osborn, T. J. Use of an upwelling-diffusion energy balance climate model to simulate and diagnose A/OGCM results. *Clim. Dyn.* **17**, 601–613 (2001).
66. Wigley, T. M. L., Ammann, C. M., Santer, B. D. & Raper, S. C. B. Effect of climate sensitivity on the response to volcanic forcing. *J. Geophys. Res.* **110**, D09107, doi:10.1029/2004JD005557 (2005).
67. Harvey, L. D. D. *et al.* in *IPCC Technical Paper 2* 1–50 (eds Houghton, J. T. *et al.*) (Intergovernmental Panel on Climate Change, Geneva, Switzerland, 1997).
68. Ammann, C. M., Meehl, G. A., Washington, W. M. & Zender, C. S. A monthly and latitudinally varying volcanic forcing dataset in simulations of 20th century climate. *Geophys. Res. Lett.* **30**, 1657, doi:10.1029/2003GL016875 (2003).
69. Lean, J. Evolution of the Sun's spectral irradiance since the Maunder Minimum. *Geophys. Res. Lett.* **27**, 2425–2428 (2000).
70. Bard, E., Raisbeck, G. M., Yiou, F. & Jouzel, J. Solar modulation of cosmogenic nuclide production over the last millennium: comparison between ¹⁴C and ¹⁰Be records. *Earth Planet. Sci. Lett.* **150**, 453–462 (1997).
71. Fisk, L. A. *et al.* Global processes that determine cosmic ray modulation. *Space Sci. Rev.* **83**, 179–214 (1998).
72. Field, C., Schmidt, G., Koch, D. & Salyk, C. Modeling production and climate-related impacts on ¹⁰Be concentration in ice cores. *J. Geophys. Res.* (in the press); preprint at (http://pubs.giss.nasa.gov/docs/notyet/2006_inpress_Field_et_al.pdf) (2006).
73. Bond, G. *et al.* Persistent solar influence on North Atlantic climate during the Holocene. *Science* **294**, 2130–2136 (2001).
74. Hu, F. S. *et al.* Cyclic variation and solar forcing of Holocene climate in the Alaskan subarctic. *Science* **301**, 1890–1893 (2003).
75. Shindell, D. T., Schmidt, G. A., Miller, R. L. & Mann, M. E. Volcanic and solar forcing of climate change during the preindustrial era. *J. Clim.* **16**, 4094–4107 (2003).
76. Harrison, R. G. The global atmospheric electrical circuit and climate. *Surv. Geophys.* **25**, 441–484, doi:10.1007/s10712-004-5439-8 (2004).
77. Haigh, J. D. The effects of solar variability on the Earth's climate. *Phil. Trans. R. Soc. A* **361**, 95–111 (2003).
78. Foukal, P. A comparison of variable solar total and ultraviolet irradiance outputs in the 20th century. *Geophys. Res. Lett.* **29**, 2089, doi:10.1029/2002GL015474 (2002).

Acknowledgements P.F. acknowledges support by NASA; T.M.L.W. was supported by the NOAA Office of Global Programs. TSI research at PMOD/WRC is supported by the Swiss National Science Foundation.

Author Contributions Author order is alphabetic. P.F. and H.S. contributed the theory of solar luminosity variation, T.M.L.W. the climate modelling and C.F. the consolidated irradiance measurements.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to H.S. (henk@mpa-garching.mpg.de).

An RNA gene expressed during cortical development evolved rapidly in humans

Katherine S. Pollard^{1*†}, Sofie R. Salama^{1,2*}, Nelle Lambert^{4,5}, Marie-Alexandra Lambot⁴, Sandra Coppens⁴, Jakob S. Pedersen¹, Sol Katzman¹, Bryan King^{1,2}, Courtney Onodera¹, Adam Siepel^{1†}, Andrew D. Kern¹, Colette Dehay^{6,7}, Haller Igel³, Manuel Ares Jr³, Pierre Vanderhaeghen⁴ & David Haussler^{1,2}

The developmental and evolutionary mechanisms behind the emergence of human-specific brain features remain largely unknown. However, the recent ability to compare our genome to that of our closest relative, the chimpanzee, provides new avenues to link genetic and phenotypic changes in the evolution of the human brain. We devised a ranking of regions in the human genome that show significant evolutionary acceleration. Here we report that the most dramatic of these 'human accelerated regions', HAR1, is part of a novel RNA gene (*HAR1F*) that is expressed specifically in Cajal-Retzius neurons in the developing human neocortex from 7 to 19 gestational weeks, a crucial period for cortical neuron specification and migration. *HAR1F* is co-expressed with reelin, a product of Cajal-Retzius neurons that is of fundamental importance in specifying the six-layer structure of the human cortex. HAR1 and the other human accelerated regions provide new candidates in the search for uniquely human biology.

The hallmark of evolutionary shift of function is sudden change in a region of the genome that previously has been highly conserved owing to negative selection. It has been speculated that changes of this type in *FOXP2*¹, a gene involved in speech production, and *ASPM*², which affects brain size, have had a significant role in the evolution of the human brain (reviewed in ref. 3). The vast majority of the approximately 15 million changes in our genome since our common ancestor with the chimpanzee are likely to represent neutral drift^{4,5}, so systematic searches for potentially important evolutionary acceleration have focused exclusively on protein coding regions^{5–9}, where there is a more favourable signal-to-noise ratio. However, protein coding regions account for only about one-third of the segments in the human genome thought to be under negative selection¹⁰, and thus these searches may be missing the majority of the functional elements in the genome^{5,11}. With the availability of nearly complete genome sequences for several vertebrates, comparative genomics can now be used to predict functional elements in the 98.5% of the genome that is non-coding, through patterns of ancestral negative selection^{10,12–16}. Here we scan these ancestrally conserved genomic regions to find those that show a significantly accelerated rate of substitution in the human lineage since divergence from our common ancestor with the chimpanzee. Many of the human accelerated regions (HARs) found in this scan are associated with genes known to be involved in transcriptional regulation and neurodevelopment. HAR1, the most dramatically changed element, is part of a novel RNA gene expressed during human cortical development.

Identification of human accelerated regions

To overcome the very low signal-to-noise ratio among changes in non-coding regions, we first searched independently for maximal-length regions of the chimpanzee genome with at least 96% identity

over 100 base pairs (bp) with the orthologous regions in mouse and rat, suggesting a significant level of negative selection in these regions. For each of the approximately 35,000 such mammalian conserved regions (median length, 140 bp), we then examined the orthologous segments in all other available amniote genomes, looking for regions that have a large number of non-adjacent changes in human relative to other species. A likelihood ratio test was used to rank and evaluate the significance of each region. Controlling the genome-wide false discovery rate (FDR) to be less than 5%, we identified 49 regions with a statistically significant substitution rate increase in human. We compared the results of this analysis to those from several complementary approaches (see Supplementary Notes S1) and found that the fastest-evolving regions score uniformly high regardless of the method (Supplementary Notes S2). Of the 49 HARs, 96% are in non-coding segments (Supplementary Table S2). Gene Ontology terms related to DNA binding and transcriptional regulation are significantly enriched among the genes adjacent to the HARs (Supplementary Table S8), and 24% of the HARs are adjacent to a neurodevelopmental gene (Supplementary Table S7), making these regions especially interesting as candidates for brain-specific regulatory elements that may have changed significantly during human evolution.

HAR1 lies in a pair of novel non-coding RNA genes

The 118-bp HAR1 region showed the most dramatically accelerated change (FDR-adjusted $P < 0.0005$), with an estimated 18 substitutions in the human lineage since the human–chimpanzee ancestor, compared with the expected 0.27 substitutions on the basis of the slow rate of change in this region in other amniotes (Supplementary Notes S3). Only two bases (out of 118) are changed between chimpanzee and chicken, indicating that the region was present

¹Center for Biomolecular Science & Engineering, ²Howard Hughes Medical Institute, and ³Center for Molecular Biology of RNA, Department of Molecular, Cell & Developmental Biology, University of California, Santa Cruz, California 95064, USA. ⁴Institut de Recherche Interdisciplinaire en Biologie Humaine et Moléculaire (IRIBHM), Free University of Brussels (ULB), and ⁵Department of Psychiatry, Erasme Hospital, Free University of Brussels (ULB), B-1070 Brussels, Belgium. ⁶INSERM, U371, Stem Cell & Brain Research, F-69500 Bron cedex, France. ⁷Université Claude Bernard, Lyon 1, F-69500 Bron cedex, France. [†]Present addresses: Department of Statistics and UC Davis Genome Center, University of California, Davis, California 95616, USA (K.S.P.); Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York 14853, USA (A.S.).

*These authors contributed equally to this work.

and functional in our ancestor at least 310 million years (Myr) ago. No orthologue of HAR1 was detected in the frog (*Xenopus tropicalis*), any of the available fish genomes (zebrafish, *Takifugu* and *Tetraodon*), or in any invertebrate lineage, indicating that it originated no more than about 400 Myr ago¹⁷. No paralogues were detected in any amniote genome draft. Resequencing in four primates further confirms that all 18 substitutions are very likely to have occurred in the human lineage (Supplementary Fig. S2). Checking against available single nucleotide polymorphisms (SNPs; <http://www.ncbi.nlm.nih.gov/SNP/>)¹⁸ and resequencing this region in a 24-person diversity panel¹⁹ indicates that all 18 substitutions are fixed in the human population. Evidence from preliminary resequencing of a 6-kilobase (kb) region containing HAR1 shows levels of polymorphism and a positive skew in the frequency spectrum that are typical of European samples²⁰, suggesting that a recent selective sweep in this region of the genome is unlikely (Supplementary Notes S4). Thus, the changes in HAR1 clearly occurred on the human lineage, but probably took place more than 1 Myr ago.

HAR1 lies in the last band of chromosome 20q and is part of a pair of overlapping divergently transcribed genes, *HAR1F* and *HAR1R*, supported by expressed sequence tag (EST) evidence. Analysis by polymerase chain reaction with reverse transcription (RT-PCR), and 5' and 3' rapid amplification of complementary DNA ends (RACE), revealed a two-exon *HAR1F* gene and an alternatively spliced *HAR1R* gene with isoforms of two and three exons (Fig. 1, Supplementary Notes S5). We sequenced the entire *HAR1F* genomic region in chimpanzee, gorilla, orang-utan and crab-eating macaque (Supplementary Notes S4, Supplementary Fig. S3). The very high divergence of the non-human primates from the human reference found in HAR1 extends through most of the first *HAR1F* exon. In addition to the primates, the dog and cow genomes also contain *HAR1F* orthologues, whereas only the HAR1 region, and not the entire transcript, can be aligned to other available amniote genomes. Most *HAR1R* regions are highly divergent, except in the chimpanzee and macaque genomes.

We have also cloned HAR1-containing transcripts from mice (Supplementary Notes S6). As in human, we found overlapping transcripts that contain HAR1. As expected by the poor conservation in the surrounding genomic region, the homology between the human and mouse HAR1 transcripts is limited to an approximately 180-bp region surrounding the HAR1 segment. Thus, although the region of overlap between human *HAR1F* and *HAR1R* (including the HAR1 segment) was conserved from our common ancestor with chicken to our common ancestor with chimpanzee, the longer *HAR1F* and *HAR1R* transcripts have evolved much more rapidly during this period.

The *HAR1F* RNA forms a stable structure

Protein coding potential of the mouse and human *HAR1F* and *HAR1R* RNAs is poor or absent (Supplementary Notes S7). However,

a phylogenetic stochastic context-free grammar model for RNA evolution¹² gives a strong prediction that the HAR1 region of the *HAR1F* transcript can fold into a stable RNA structure (Fig. 2), and a slightly less-strong prediction for the same region of the *HAR1R* transcript (data not shown). Three of five RNA helices in the predicted secondary structure are statistically confirmed by five pairs of human-specific compensatory changes that preserve Watson–Crick pairing: four pairs of transitions and one pair of transversions ($P = 3.95 \times 10^{-8}$; Supplementary Notes S8). A different compensatory pair of transversions is also found in the platypus genome. No changes inconsistent with the core of any helix, excluding the terminal base pairs and the second pair of the outermost enclosing helix in mouse, were found in any sequence match from the National Center for Biotechnology Information (NCBI) trace repository (one match from each of 11 amniote species; Supplementary Fig. S2).

We tested this secondary structure prediction by performing structure probing experiments using dimethyl sulphate (DMS) treatment and primer extension of *in vitro* synthesized human and chimpanzee *HAR1F* RNAs (Supplementary Notes S9)²¹. The results were largely consistent with the predicted structure, but we observed several differences between the human and chimpanzee RNAs (Fig. 2c and Supplementary Fig. S10). In particular, our results are consistent with a human-specific structure where helix D is extended at the expense of helix C (compare Fig. 2b to Fig. 2c). In contrast, excluding the human sequence from the comparative analysis leads to an evolutionarily conserved structure with three additional base pairs in helix C and two fewer pairs in helix D. This structure agrees with the protection of bases in helix C and the accessibility of bases in the shorter helix D we observe for the chimpanzee RNA. In the human RNA, bases in helix D are strongly protected whereas those in helix C are more accessible by DMS. These results indicate that although selective forces to preserve base pairing interactions shaped this region, the structure of the *HAR1F* RNA may be significantly different between humans and their ancestors, and may even lack helices C and E. This RNA structure is novel; no primary or secondary structural homology was found with any RNA in the Rfam RNA repository²², and no convincing evidence of microRNA (miRNA) precursors in these transcripts has yet been found (Supplementary Notes S10). On the basis of this analysis, it seems that *HAR1F* and *HAR1R* are novel non-coding RNA genes.

HAR1F is expressed in the developing human and primate brain

To determine if *HAR1F* is expressed in the developing human central nervous system, we performed RNA *in situ* hybridizations on human embryonic brain sections. We found that *HAR1F* is strongly and specifically expressed in the developing neocortex early in human embryonic development (Fig. 3). *HAR1F* was first detected at 7 and 9 gestational weeks (GW), specifically in the dorsal telencephalon (the primordium of the cerebral cortex) and not detected in other parts of

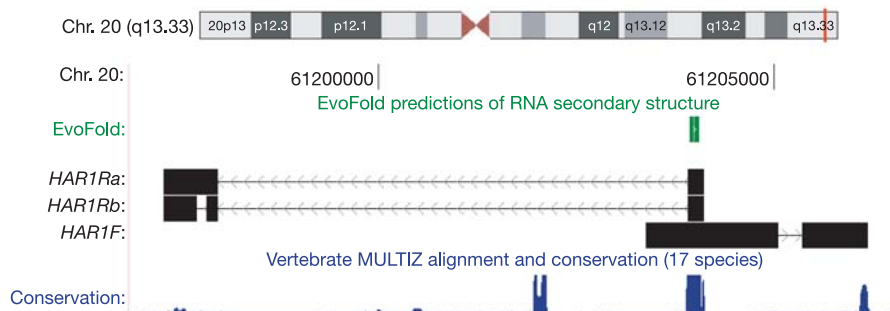


Figure 1 | HAR1-associated transcripts in genomic context. Schematic drawing showing the genomic context on chromosome 20q13.33 of the HAR1-associated transcripts *HAR1F* and *HAR1R* (black, with a chevroned line indicating introns), and the predicted RNA structure (green) based on

the May 2004 human assembly in the UCSC Genome Browser⁴¹. The level of conservation in the orthologous region in other vertebrate species (blue) is plotted for this region using the PhastCons program¹⁶. Both the common and testes-specific splice sites are conserved (data not shown).

the forebrain (Fig. 3a and data not shown). Within the early developing cortex, *HARIF* was selectively expressed in a subset of cells located close to the pial surface in the marginal zone (Fig. 3a, b), indicating its expression in Cajal–Retzius neurons. This was confirmed by co-detection of reelin protein, a specific marker of these neurons, and *HARIF* RNA (Fig. 3a, c, d), which showed that both *HARIF* and reelin are expressed in the same cells. In contrast, *HARIF* did not seem to be expressed in interneurons expressing vesicular GABA (γ -aminobutyric acid) transporter (vGAT), the other main neuronal population of the marginal zone (Fig. 3d). The expression of *HAR1* in reelin-positive neurons was maintained until at least 17–19 GW (Fig. 3c, d), at which stage reelin-positive Cajal–Retzius-like neurons are interspersed within the subpial granular layer (SGL)—a layer thought to contain Cajal–Retzius neurons and their precursors

migrating tangentially from outside the neocortex^{23–25}, and which is uniquely developed in humans^{24–27}. *HARIF* was also expressed in cells of the upper cortical plate (Fig. 3b, c), presumably corresponding to neurons finishing their radial migration. Notably, the human reelin receptor gene, *VLDLR*, also shares this expression pattern in humans, but is not expressed in Cajal–Retzius neurons in the mouse²⁸. At later stages (24 GW), expression of *HARIF* was no longer detectable in Cajal–Retzius cells. At 17–24 GW, expression of *HARIF* was observed in brain regions other than the cortex: in particular, the hippocampal primordium, the dentate gyrus, the developing cerebellar cortex, and a few hindbrain nuclei such as the olivary complex (Fig. 4). In contrast to the distinct pattern of *HARIF* expression, probes specific for *HAR1R* show no specific staining in these embryonic sections (data not shown).

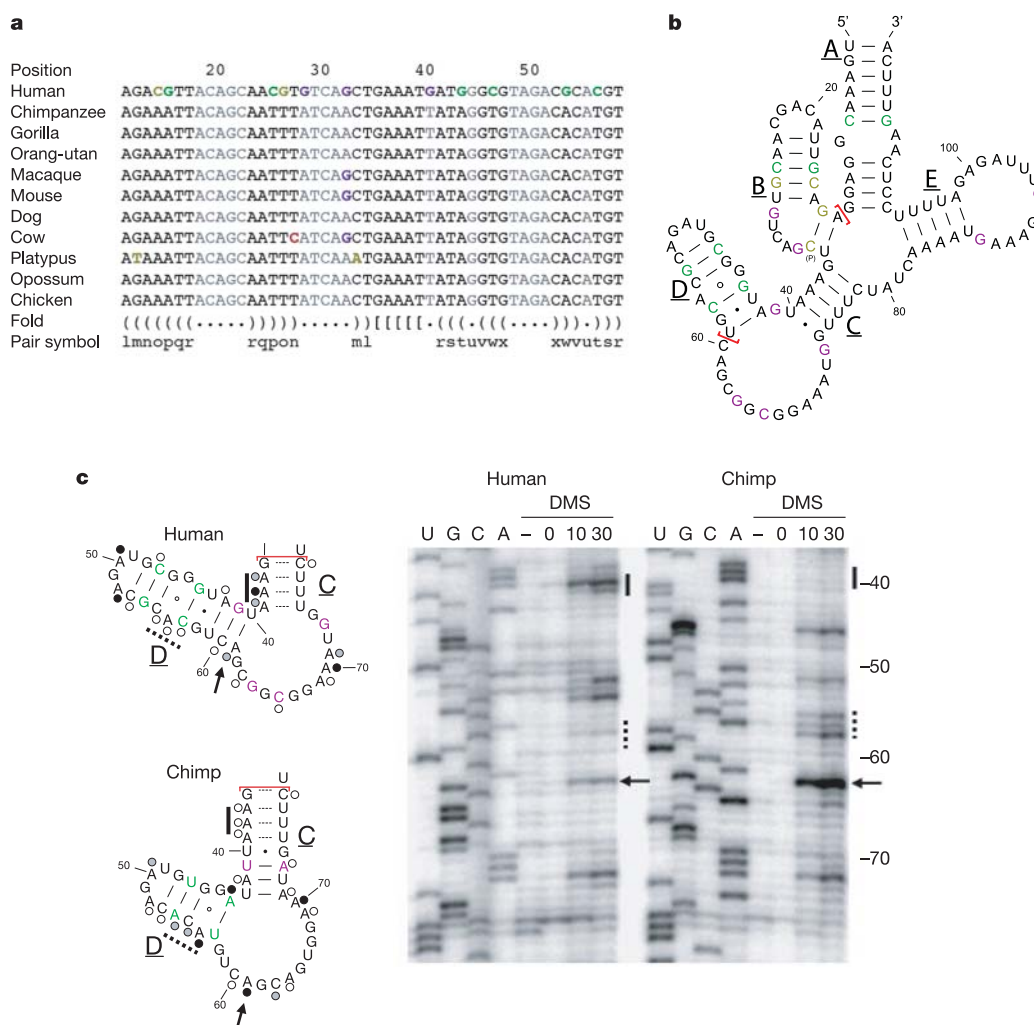


Figure 2 | Predicted RNA secondary structure for *HARIF*. **a**, Section of the multiple alignment of *HARIF* in various amniote species. The multiple alignment is annotated with the secondary structure (fold) shown in panel **b**. Matching round parentheses indicates pairing bases. Square parentheses are used to indicate bases that are predicted to pair outside the region shown. Unpaired regions are shown in grey and substitutions are shown with the following colour scheme: green denotes compensatory transitions, yellow–green denotes compensatory transversions, purple denotes substitutions in unpaired regions, and red denotes non-compensated changes. **b**, The evolutionarily conserved parts of the RNA secondary structure of the *HARIF* region as predicted using the EvoFold program¹². Substitutions are shown using the colour scheme in panel **a**; red bars indicate the region for which the alignment is shown in panel **a**. The structure is supported by substitutions on the human lineage, as well as a pair of changes in platypus (indicated by P)). Only the helices with the compensatory substitutions can

be considered to be supported by evolutionary data. **c**, Distinct secondary structure predictions for helices C and D in the human sequence (Human) versus non-human sequences (Chimp), on the basis of combined comparative sequence analysis and structure probing. Circles indicate the bases that can be modified by DMS and indicate strong (black), weak (grey) or no (white) modification as shown in the gel to the right. For each species, lanes 1–4 show the sequencing reactions labelled according to the corresponding base in the RNA structure; lanes 5–8 show the primer extension reactions of RNA after no (–), 0, 10 or 30 min of DMS treatment at 18 °C. Numbers on the right indicate the position in RNA structure; arrows indicate the modified base at position 61; solid lines indicate the position of bases 37–39; and dashed lines indicate the position of bases 54–57. The presence of a band in the last two lanes indicates modification of the base at the previous position (that is, the arrows at position 62 on the gel indicate a modified A at position 61).

We examined the evolutionary conservation of this expression pattern by performing RNA *in situ* hybridizations on brain sections of another primate, the cynomolgus macaque. For these experiments, we cloned a fragment of macaque DNA corresponding to most of *HAR1F* exon 1 and tested it on embryonic day (E) 79–85 samples, corresponding to midgestation and to stages of active cortical neuron production and radial migration²⁹. The *HAR1F* expression pattern was very similar to that observed in the human sections (Fig. 3e). *HAR1F* is co-expressed with reelin in the SGL and is also expressed in cells of the upper cortical plate. Thus, despite extensive sequence changes, the expression pattern of *HAR1F* in the developing cortex has been highly conserved since the divergence of hominoids and Old World monkeys some 25 Myr ago. This suggests that the *HAR1F* expression pattern is functionally significant and is reminiscent of the evolution of *ASPM*,

which shows rapid change in its coding sequence but conserved patterns of expression in neural progenitors^{2,30}.

HAR1 genes are expressed in the adult brain

We also examined *HAR1F* and *HAR1R* expression in adult human samples by using tissue RNA preparations for northern blot and quantitative real-time PCR (qPCR) analysis (Table 1), as well as RNA *in situ* hybridization. *In situ* hybridization experiments on adult brain sections revealed a diffuse pattern of staining for *HAR1F* in the frontal cortex and hippocampus (Supplementary Fig. S12). Although no specific signal was detected for either transcript by northern blot analysis (data not shown), we detected expression of both transcripts by qPCR in total brain, ovary and testes samples. Neither transcript was detected in other tissues tested, including adrenal, bladder,

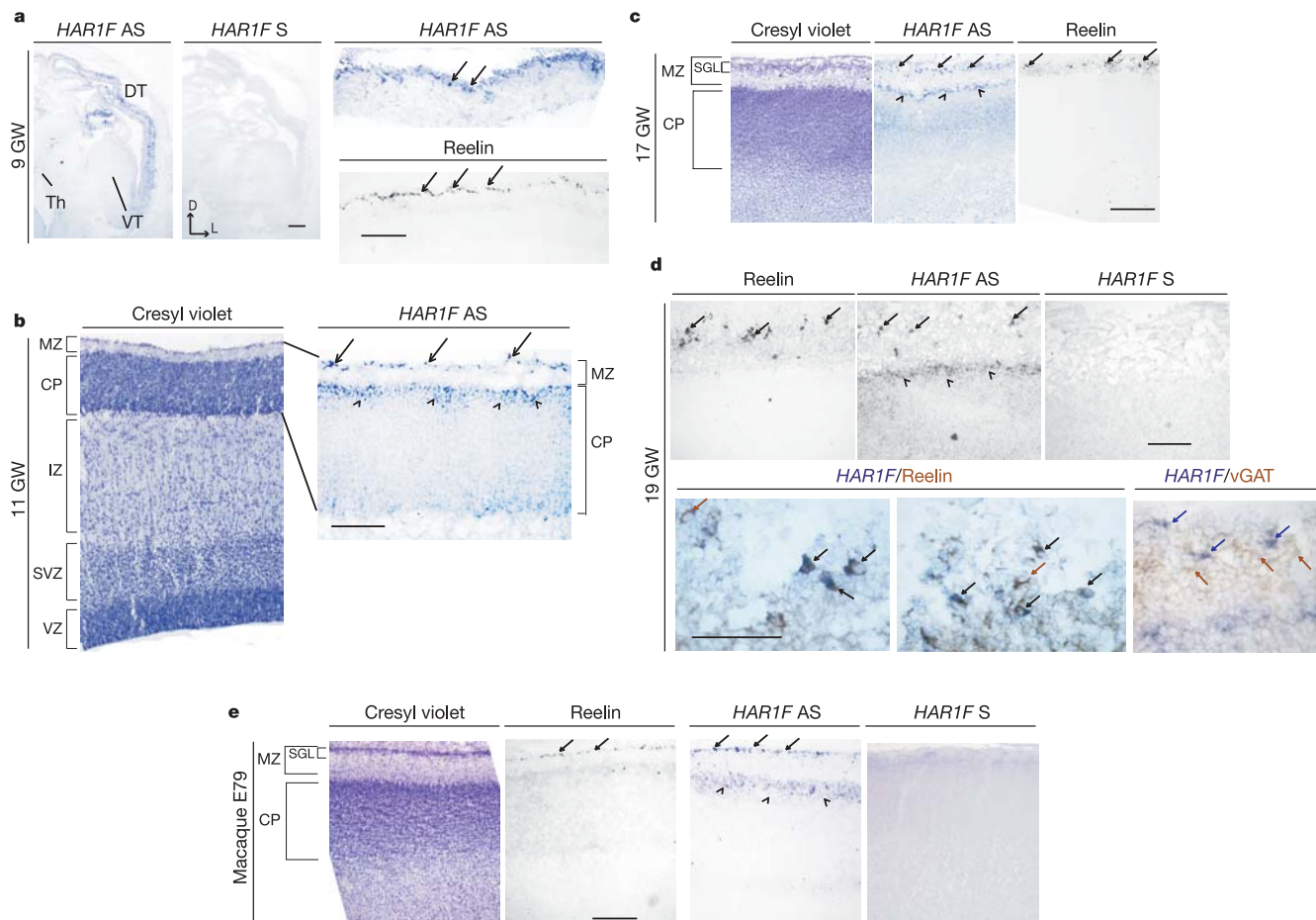


Figure 3 | Expression of *HAR1F* and *HAR1R* in the developing neocortex. Pattern of expression of *HAR1F* and reelin in the human and macaque developing brain, revealed by RNA *in situ* hybridization (*HAR1F*) and immunohistochemistry (reelin). **a**, *In situ* hybridization of *HAR1F* antisense (AS) and sense (S) probes on 9-GW human brain coronal sections, illustrating selective expression of *HAR1F* in the dorsal telencephalon (DT), whereas the ventral telencephalon (VT) and thalamus (Th) show no obvious expression. The *HAR1* probes contain 112 bp of the region of overlap between *HAR1F* and *HAR1R*. Therefore, the *HAR1F* sense probe has the potential to detect *HAR1R* expression. Within the DT, *HAR1F* is selectively enriched in the most superficial part of the cortex, much like reelin at the same stage (arrows). Scale bars: 500 μ m on left panels; 250 μ m on right panels. Dorsal (D) is top; lateral (L) is right on left panels. **b**, At 11 GW, *HAR1F* expression is confined to the cerebral cortex, with expression in large cells situated at the border of the marginal zone (MZ) (arrows), corresponding to Cajal–Retzius neurons, and in the upper cortical plate (CP) (arrowheads). Scale bar: 125 μ m for the right panel. IZ, intermediate zone; SVZ, subventricular zone; VZ, ventricular zone. **c**, At 17 GW, expression of *HAR1F* and reelin remain confined to Cajal–Retzius cells in

the MZ (arrows), intermingled with other cells of the subplial granular layer (SGL). In addition, *HAR1F* is expressed in neurons of the upper CP (arrowheads). Scale bar: 1,000 μ m. **d**, At 19 GW, *HAR1F* and reelin show the same pattern in the MZ/SGL (arrows), and *HAR1F* remains expressed in the upper CP (arrowheads), whereas the *HAR1F* sense probe shows no signal. Lower panels centred on the MZ illustrate double *in situ* hybridization/immunohistochemistry experiments. Lower left and middle panels show that *HAR1F* (in dark blue) and reelin (in dark brown) are co-localized in most cells, corresponding to Cajal–Retzius neurons (black arrows), although some reelin-positive cells do not express *HAR1F* (brown arrows). The lower right panel illustrates that *HAR1F* cells (blue arrows) seem to be mainly distinct from interneurons (brown arrows) expressing vesicular GABA transporter (vGAT). Scale bar: 250 μ m. **e**, Expression of *HAR1F* in macaque developing cortex. *In situ* hybridization (*HAR1F*) and immunohistochemistry (reelin) on E79 macaque brain parasagittal sections. As in the human, *HAR1F* and reelin are expressed in Cajal–Retzius cells in the MZ (arrows), intermingled with other cells of the SGL. In addition, *HAR1F* is expressed in neurons of the upper CP (arrowheads). Scale bar: 1,000 μ m. The *HAR1F* sense probe does not reveal any obvious expression pattern.

breast, colon, liver, pancreas, placenta, skeletal muscle and thymus. To further characterize the expression in brain, we obtained RNA samples from specific adult human brain regions as well as a fetal brain RNA sample. Expression of *HAR1F* is highest in cerebellum and is also prominent in forebrain structures, including the cortex, hippocampus, thalamus and hypothalamus. In general, *HAR1F* expression is higher than *HAR1R*. Testis is an exception; here transcripts are present in small, but roughly equal, amounts. It is interesting to note that *HAR1R* levels are dramatically lower than *HAR1F* (50-fold lower) in the fetal brain sample, consistent with our inability to detect it with *in situ* hybridization. The finding that both transcripts overlap and are expressed in the same tissues raises the possibility of antisense regulation between them. So far, however, we have not detected expression of the two transcripts in the same cells.

In a qPCR analysis of mouse *HAR1F* and *HAR1R* expression (Supplementary Notes S6), we found that both *HAR1F* and *HAR1R* are expressed in embryonic development, and the peak expression of both *HAR1F* and *HAR1R* was at E15. In contrast to our observations in human fetal brain samples, *HAR1F* expression was not significantly higher than *HAR1R* expression. In adult tissues, we found both *HAR1F* and *HAR1R* specifically expressed in brain regions, but again the levels of *HAR1F* and *HAR1R* were similar. These results suggest that the regulation of the *HAR1* region transcripts may have changed since the common ancestor of mouse and human.

HAR1 substitutions show a weak-to-strong bias

A curious feature of the evolution of *HAR1* is that all 18 human-specific changes in the predicted RNA structure are from A or T ('W') nucleotides to G or C ('S') nucleotides. This has the effect of strengthening RNA helices in the predicted RNA structure. However, this pattern of highly biased substitutions extends well beyond the predicted RNA structure in a 1.2-kb region overlapping the first exon of *HAR1F* (34 W → S substitutions out of 44 total human–chimpanzee differences, with no S → W substitutions). Notably, a W → S substitution bias is present in many of the human accelerated regions. Among the 49 most accelerated, the frequency of W → S substitutions is 77% higher than that of S → W substitutions, even after a correction for variation in the overall base composition of the ancestral sequence, whereas there is no difference in the larger set

of 35,000 conserved regions (Supplementary Notes S11). Compared with an expected <5%, we find that 12% of these 49 regions, including *HAR1*, lie in the final band of their chromosome arms, where there is an increased rate of recombination¹⁴. Recombination hotspots are known to differ in location between human and chimpanzee³¹, and have been speculated to be associated with increases in W → S substitutions through a process known as biased gene conversion³². Selection for increased gene stability or increased expression levels have been proposed as alternative explanations for W → S substitution biases^{33,34}. Bias towards W → S substitution through any of these general processes may have combined with selection for specific compensatory changes in the RNA structure to dramatically alter *HAR1* during human evolution.

Conclusion

One primary issue remaining concerns the role of *HAR1F* and *HAR1R* in cortical development. Although *HAR1F* is co-expressed with *reelin*—a known regulator of cortical development—within Cajal–Retzius cells of the SGL, it remains to be determined whether *HAR1F* influences the expression of *reelin* or its receptors either directly or indirectly, or whether it acts through a different pathway. Notably, the extra-cortical sites of expression of *HAR1F* (such as cerebellar cortex and olivary nuclei) also correspond to structures that are patterned by the *reelin* pathway in mouse and human³⁵. A suggested association between *reelin*-expression-related defects in the brain and schizophrenia provides additional motivation for further investigation of this issue³⁶. Additionally, the pattern of expression of *HAR1R* suggests, but does not prove, that it may be expressed later in development to downregulate *HAR1F* by antisense inhibition, as is the case for some other developmental genes³⁷.

Our comparative genomic approach to identify the most dramatically changed segments of DNA in the human genome has identified a number of new candidate regions to test for clues in the attempt to decode the key events in human evolution. The first on this list seems to define a new type of RNA molecule expressed at a very crucial time and place in the development of the neocortex. Many of the other top candidates are associated with genes known to be involved in neurodevelopment, an area where there has been significant divergence since our last common ancestor with chimpanzee. Thus, this seems to be a promising approach to identifying candidate regions involved in neurodevelopmental aspects of our uniquely human biology.

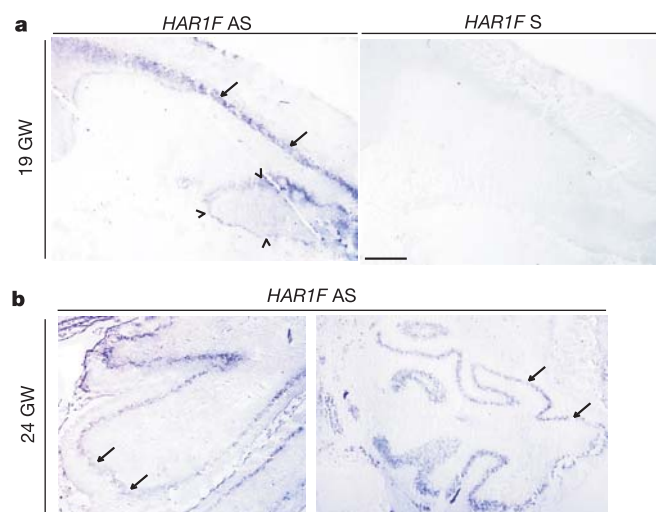


Figure 4 | Expression of *HAR1F* in other parts of the developing brain. Pattern of expression of *HAR1F* in the developing brain, revealed by RNA *in situ* hybridization. **a**, *In situ* hybridization on 19-GW coronal brain sections, illustrating expression of *HAR1F* in the hippocampal primordium (arrows) and dentate gyrus (arrowheads). *HAR1F* sense probe detects no obvious signal. Scale bar: 750 μ m. **b**, *In situ* hybridization on 24-GW coronal brain sections, illustrating expression of *HAR1F* in the cerebellar cortex (arrows in left panel) and olivary complex (arrows in right panel).

Table 1 | Expression of human *HAR1F* and *HAR1R*

Sample	<i>HAR1F</i> *	<i>HAR1R</i> *
Total RNA		
Cerebral cortex	1.0† (0.77, 1.3)	0.095 (0.054, 0.17)
Cerebellum	4.1 (2.4, 7.2)	0.041 (0.030, 0.055)
Frontal lobe	1.2 (0.70, 2.0)	0.12 (0.067, 0.22)
Temporal lobe	0.72 (0.55, 0.95)	0.049 (0.029, 0.083)
Parietal lobe	0.77 (0.61, 0.98)	0.083 (0.052, 0.13)
Occipital pole	1.1 (0.92, 1.3)	0.12 (0.065, 0.21)
Insula	0.91 (0.62, 1.3)	0.078 (0.045, 0.13)
Hippocampus	0.65 (0.44, 0.96)	0.051 (0.031, 0.087)
Pons	0.51 (0.35, 0.76)	0.21 (0.11, 0.40)
Medulla oblongata	0.39 (0.27, 0.56)	0.043 (0.025, 0.074)
Fetal brain	0.14 (0.11, 0.18)	0.003 (0.002, 0.005)
Brain	0.96 (0.71, 1.3)	0.024 (0.015, 0.039)
Testes	0.031 (0.006, 0.16)	0.12 (0.047, 0.31)
Ovary	0.19 (0.15, 0.25)	0.009 (0.006, 0.012)
PolyA ⁺ RNA‡		
Thalamus	4.9 (3.2, 7.5)	0.51 (0.27, 0.98)
Hypothalamus	4.8 (3.5, 6.6)	0.54 (0.30, 0.96)

* Values are expressed as a ratio of the expression in the cerebral cortex sample of *HAR1F*. Entries represent the averages and 95% confidence intervals (in parentheses) of six replicates from two independent experiments, with the exception of brain *HAR1R*, which represents nine replicates from two independent experiments.

† A value of 1.0 is equivalent to 24,014 copies in 100 ng of template cDNA.

‡ Approximately one-third the amount of template was used for polyA⁺-derived samples. It is expected that expression will be higher in polyA⁺-enriched samples than in samples derived from total RNA.

METHODS

Detailed methods are given in Supplementary Information. Using alignments produced by MULTIZ (http://www.bx.psu.edu/miller_lab/), we identified 34,498 conserved regions of the chimpanzee genome that are ≥ 100 bp long and $\geq 96\%$ identical with mouse and rat. Each conserved region was evaluated for acceleration in the human lineage using a likelihood ratio test (LRT) using DNA from other species. Statistical significance was assessed by simulation and *P*-values are adjusted for multiple comparisons.

For RT-PCR, RACE and qPCR analysis, RNAs (Ambion and Clontech) were used to synthesize cDNA primed with a combination of oligo-dT and random hexamer primers using Superscript III (Invitrogen) according to the manufacturer's recommendations.

The qPCR analysis used iQSYBR Green Supermix (BioRad) with $2.5 \text{ ng } \mu\text{L}^{-1}$ (or $0.8 \text{ ng } \mu\text{L}^{-1}$ for polyA⁺-derived samples) cDNA template in an i-Cycler thermal cycler (BioRad); RACE analysis used the GeneRacer kit (Invitrogen). Secondary structure models were based on an EvoFold prediction¹², and chemical probing of unpaired A and C residues with dimethyl sulphate (DMS) was carried out using methods described previously²¹.

RNA *in situ* hybridization using digoxigenin-labelled RNA probes adapted to human brain specimens was performed as described previously³⁸. Immunodetection was performed as described³⁹ using mouse monoclonal anti-reelin antibody 142 (ref. 40; kindly provided by A. Goffinet) and a rabbit antibody to vGAT (Synaptic Systems). Between 2 and 16 sections from each case were tested.

This study was approved by the three relevant local ethics committees (Erasmus Academic Hospital, University of Brussels, and Belgian National Fund for Scientific Research) on research involving human subjects. Written informed consent was given by the parents in each case. All experiments were performed in compliance with the national and European laws, as well as with INSERM institutional guidelines, concerning animal experimentation.

Received 27 June; accepted 25 July 2006.

Published online 16 August 2006.

- Enard, W. *et al.* Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).
- Evans, P. D. *et al.* Adaptive evolution of *ASPM*, a major determinant of cerebral cortical size in humans. *Hum. Mol. Genet.* **13**, 489–494 (2004).
- Hill, R. S. & Walsh, C. A. Molecular insights into human brain evolution. *Nature* **437**, 64–67 (2005).
- Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
- Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
- Clark, A. G. *et al.* Inferring nonneutral evolution from human–chimp–mouse orthologous gene trios. *Science* **302**, 1960–1963 (2003).
- Nielsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, e170 (2005).
- Dorus, S. *et al.* Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell* **119**, 1027–1040 (2004).
- Wang, X., Grus, W. E. & Zhang, J. Gene losses during human origins. *PLoS Biol.* **4**, e52 (2006).
- Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
- Pedersen, J. S. *et al.* Identification and classification of conserved RNA secondary structures in the human genome. *PLOS Computat. Biol.* **2**, e33 (2006).
- Hillier, L. W. *et al.* Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
- Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning human gene deserts for long-range enhancers. *Science* **302**, 413 (2003).
- Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
- Hedges, S. B. & Kumar, S. Genomics. Vertebrate genomes compared. *Science* **297**, 1283–1285 (2002).
- Altshuler, D. *et al.* A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
- Collins, F. S., Brooks, L. D. & Chakravarti, A. A. DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* **8**, 1229–1231 (1998).
- Stajich, J. E. & Hahn, M. W. Disentangling the effects of demography and selection in human history. *Mol. Biol. Evol.* **22**, 63–73 (2005).
- Merryman, C. & Noller, H. in *RNA–Protein Interactions: A Practical Approach* (ed. Smith, C.) 237–253 (Oxford Univ. Press, New York, 1998).
- Griffiths-Jones, S. *et al.* Rfam: Annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124 (2005).
- Meyer, G. & Goffinet, A. M. Prenatal development of reelin-immunoreactive neurons in the human neocortex. *J. Comp. Neurol.* **397**, 29–40 (1998).
- Meyer, G. & Wahle, P. The paleocortical ventricle is the origin of reelin-expressing neurons in the marginal zone of the foetal human neocortex. *Eur. J. Neurosci.* **11**, 3937–3944 (1999).
- Zecevic, N. & Rakic, P. Development of layer I neurons in the primate cerebral cortex. *J. Neurosci.* **21**, 5607–5619 (2001).
- Meyer, G., Soria, J. M., Martinez-Galan, J. R., Martin-Clemente, B. & Fairen, A. Different origins and developmental histories of transient neurons in the marginal zone of the fetal and neonatal rat cortex. *J. Comp. Neurol.* **397**, 493–518 (1998).
- Rakic, S. & Zecevic, N. Emerging complexity of layer I in human cerebral cortex. *Cereb. Cortex* **13**, 1072–1083 (2003).
- Perez-Garcia, C. G., Tissir, F., Goffinet, A. M. & Meyer, G. Reelin receptors in developing laminated brain structures of mouse and human. *Eur. J. Neurosci.* **20**, 2827–2832 (2004).
- Lukaszewicz, A. *et al.* G1 phase regulation, area-specific cell cycle control, and cytoarchitectonics in the primate cortex. *Neuron* **47**, 353–364 (2005).
- Bond, J. & Woods, C. G. Cytoskeletal genes regulating brain size. *Curr. Opin. Cell Biol.* **18**, 95–101 (2006).
- Ptak, S. E. *et al.* Fine-scale recombination patterns differ between chimpanzees and humans. *Nature Genet.* **37**, 429–434 (2005).
- Duret, L., Semon, M., Piganeau, G., Mouchiroud, D. & Galtier, N. Vanishing GC-rich isochores in mammalian genomes. *Genetics* **162**, 1837–1847 (2002).
- Bernardi, G. Isochores and the evolutionary genomics of vertebrates. *Gene* **241**, 3–17 (2000).
- Kudla, G., Lipinski, L., Caffin, F., Helwak, A. & Zylicz, M. High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biol.* **4**, e180 (2006).
- Tissir, F. & Goffinet, A. M. Reelin and brain development. *Nature Rev. Neurosci.* **4**, 496–505 (2003).
- Impagnatiello, F. *et al.* A decrease of reelin expression as a putative vulnerability factor in schizophrenia. *Proc. Natl Acad. Sci. USA* **95**, 15718–15723 (1998).
- Alfano, G. *et al.* Natural antisense transcripts associated with genes involved in eye development. *Hum. Mol. Genet.* **14**, 913–923 (2005).
- Lambot, M. A., Depasse, F., Noel, J. C. & Vanderhaeghen, P. Mapping labels in the human developing visual system and the evolution of binocular vision. *J. Neurosci.* **25**, 7232–7237 (2005).
- Depaepe, V. *et al.* Ephrin signalling controls brain size by regulating apoptosis of neural progenitors. *Nature* **435**, 1244–1250 (2005).
- de Bergeyck, V., Naerhuyzen, B., Goffinet, A. M. & Lambert de Rouvroit, C. A panel of monoclonal antibodies against reelin, the extracellular matrix protein defective in reeler mutant mice. *J. Neurosci. Methods* **82**, 17–24 (1998).
- Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Bejerano, C. Lowe, J. Kent, H. Noller, D. Feldheim, A. Love (UCSC), V. Albert and J.-C. Noel (Erasmus Hospital), J.-P. Brion (ULB), A. Goffinet (UCL Louvain), the UCSC Genome Browser Group, Webb Miller (Penn. State), the Macaque Genome Sequencing Consortium, and the Broad Institute Genome Sequencing Platform and Whole-Genome Assembly Team. This work was funded by the Howard Hughes Medical Institute (D.H., S.R.S. and B.K.), the US NHGRI (D.H., A.D.K., S.K. and C.O.), the US National Cancer Institute (J.S.P.), the US NIGMS (K.S.P. and M.A.), the University of California Biotechnology Research and Education Program (A.S.), the Danish Research Council (J.S.P.), the Belgian FNRS, the Belgian FRSM, the Belgian Queen Elizabeth Medical Foundation (FMRE), the Interuniversity Attraction Poles Programme, Belgian State and Federal Office for Scientific, Technical and Cultural Affairs (P.V.), and the Fondation Erasme (M.-A.L. and N.L.). P.V. and M.-A.L. are Research Associate and Research Fellow of the FNRS, respectively.

Author Contributions Computational methods developed and applied by K.S.P., J.S.P., A.S. and A.D.K. Experimental analysis by S.R.S., P.V., N.L., M.-A.L., S.C., S.K., B.K., C.O., C.D., H.I. and M.A. Manuscript primarily written by K.S.P., S.R.S., P.V. and D.H.

Author Information The sequences of the HARI transcripts reported in this study have been submitted to GenBank (accession numbers DQ860409–DQ860415). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.H. (haussler@soe.ucsc.edu).

Crystal structures of a multidrug transporter reveal a functionally rotating mechanism

Satoshi Murakami^{1,2,3,4,6}, Ryosuke Nakashima¹, Eiki Yamashita⁵, Takashi Matsumoto^{1,3} & Akihito Yamaguchi^{1,3,4}

AcrB is a principal multidrug efflux transporter in *Escherichia coli* that cooperates with an outer-membrane channel, TolC, and a membrane-fusion protein, AcrA. Here we describe crystal structures of AcrB with and without substrates. The AcrB–drug complex consists of three protomers, each of which has a different conformation corresponding to one of the three functional states of the transport cycle. Bound substrate was found in the periplasmic domain of one of the three protomers. The voluminous binding pocket is aromatic and allows multi-site binding. The structures indicate that drugs are exported by a three-step functionally rotating mechanism in which substrates undergo ordered binding change.

Multidrug resistance caused by export proteins is a serious problem, not only in the chemotherapy of cancer but also in the antibiotic treatment of numerous different bacterial infections¹. AcrB is a principal multidrug exporter^{2,3} expressed almost constitutively and confers intrinsic drug tolerance to *E. coli*^{2–6}. AcrB belongs to the resistance–nodulation–division (RND) family of transporters⁷, and its homologues occur in many pathogenic species of Gram-negative bacteria⁸. It cooperates with both an outer-membrane channel, TolC, and a membrane-fusion protein, AcrA^{9,10}, and exports a wide variety of toxic compounds—including anionic, cationic, zwitterionic and neutral compounds⁵—directly out of the cell by bypassing the periplasm^{11,12}.

Our previous crystal structures of AcrB showed that it exists as a trimer, consisting of three layers parallel to the membrane: a transmembrane domain, a porter domain (formerly named the “pore” domain), and a TolC docking domain (Fig. 1a)^{13,14}. In the porter domain, there is a functionally important helix bundle—the central helices—composed of three α -helices from each protomer (Fig. 1b)¹⁵. In the TolC docking domain, AcrB opens like a funnel that fits to the proximal end of TolC, indicating the direct docking of AcrB and TolC¹⁶. TolC is also a trimer, and forms a tube extending across the periplasm and the outer membrane to provide an export pathway to the outside of the cell¹⁷. The structures indicate that AcrB takes substrates from the periplasm through the opening at the membrane–periplasm boundary between each protomer, and extrudes them from the top funnel into the TolC channel¹⁵. The structures of the membrane-fusion protein AcrA, and its homologue MexA from *Pseudomonas aeruginosa*, suggest that the AcrB–TolC complex is surrounded by AcrA^{18–20}.

In the AcrA–AcrB–TolC tripartite complex, AcrB determines the substrate specificity and actively exports a wide variety of structurally dissimilar drugs and toxic compounds using the proton motive force. The questions that remain are how the multidrug efflux transporter recognizes such a wide variety of toxic compounds, and how it actively exports them out of the cell. In order to address these questions, we needed to determine the structure of the AcrB–

substrate complex. Such attempts have been made previously, but there was a possible artefact arising from the crystal symmetry²¹.

We now have AcrB crystals belonging to the C2 space group, which allows each protomer to take a different conformation. Three protomers that constitute a trimer are now clearly differentiated in the electron density map at 2.8 Å resolution (Supplementary Fig. S1), and only one of them has a bound substrate in a phenylalanine-rich pocket in the periplasmic region²². We propose that the three different conformations of the protomers represent the three states of drug export: access, binding and extrusion. Presumably, the drugs are exported by a functionally rotating mechanism for ordered binding change.

Structure determination

The C2 crystals of substrate-free AcrB and AcrB–substrate complex were generated (see Methods). The structure of the unliganded form was solved at 2.8 Å resolution by the multiple isomorphous replacement method (Supplementary Table S1). Bound substrates were located by difference Fourier calculation, and the bromine atom of the brominated substrate by anomalous difference Fourier map (Fig. 2a).

Overall structure of the asymmetric AcrB–drug complex

Our new crystal structure solved with a C2 crystal is basically consistent with our previous structure of a trigonal R32 crystal, having crystallographic three-fold symmetry, except for some crucial differences in the conformation of each protomer. The root-mean-square deviation value of the superposition of main-chain traces of these two structures is ~ 2.4 Å. The substrate is bound to only one of the three protomers, named the ‘binding’ protomer (Fig. 1, blue). No substrate is found in the central cavity, even at the maximally dissolvable concentration of drugs for co-crystallization.

The other conspicuous difference is observed in the triplet of central helices in the porter domain. One of the three helices (Fig. 1, red) is inclined nearly 15° towards the binding protomer, compared with the other two helices. The potential exit between the distal part

¹Department of Cell Membrane Biology, Institute of Scientific and Industrial Research, Osaka University, Ibaraki, Osaka 567-0047, Japan. ²PRESTO, and ³CREST, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan. ⁴Faculty of Pharmaceutical Science, and ⁵Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan. ⁶SOSHO Inc., Hommachi, Osaka 541-0053, Japan.

of PN1 and PN2 (subdomains within the porter domain) from the substrate binding pocket in the binding protomer is blocked with this inclined α -helix (Fig. 1d and Fig. 2). In contrast, the exit from the vacant binding site of the protomer contributing this inclined helix is open towards the exit funnel. We suggest that this protomer represents the situation just after extrusion of the substrates, and named it the 'extrusion' protomer (Fig. 1 and Fig. 3a, red). The third protomer has a vacant binding site but an upright central helix, representing an intermediate state between extrusion and binding; thus, it is named the "access" protomer, waiting for the next substrate binding event (Fig. 1, green).

Drug binding site

A substrate binding pocket exists in the porter domain of each protomer. The porter domain consists of four subdomains—PN1, PN2, PC1 and PC2¹³—each of which consists of two β - α - β sandwiches. These four subdomains are packed with their β -sheets surrounding a voluminous pocket (Fig. 1d). Bound minocycline and doxorubicin are observed in one of the three protomers at the

distal end of the pocket, between the β -sheets of PN2 and PC1 (Fig. 1d and Fig. 2). The bromine atom of the 9-bromo-minocycline was clearly identified in the anomalous difference Fourier map (6.2σ ; $0.02417 \text{ electrons } \text{\AA}^{-3}$) at the expected position for the structure of minocycline (Fig. 2a). The substrate binding pocket is rich in aromatic amino-acid residues: Phe 136 and Phe 178 (PN2), and Phe 610, Phe 615, Phe 617 and Phe 628 (PC1). These aromatic side chains may interact with the drug molecule by hydrophobic or aromatic–aromatic interactions (Fig. 1d). There are some polar residues in this area, such as Asn 274 and Gln 176, possibly forming hydrogen bonds with the drug molecule.

Figure 2b, c shows details of the binding site with substrates bound. The methyl moiety of the 7-dimethylamino group and the C-ring of minocycline interact with Phe 178 and Phe 615, respectively. Two oxygen atoms of the 1-oxo and 2-amido groups interact with Asn 274 (see also Supplementary Fig. S2). Doxorubicin interacts with Phe 615, similar to minocycline, but interacts with Gln 176 and Phe 617 instead of Asn 274 and Phe 178, respectively. Therefore, different sets of residues are used for binding of the different kinds

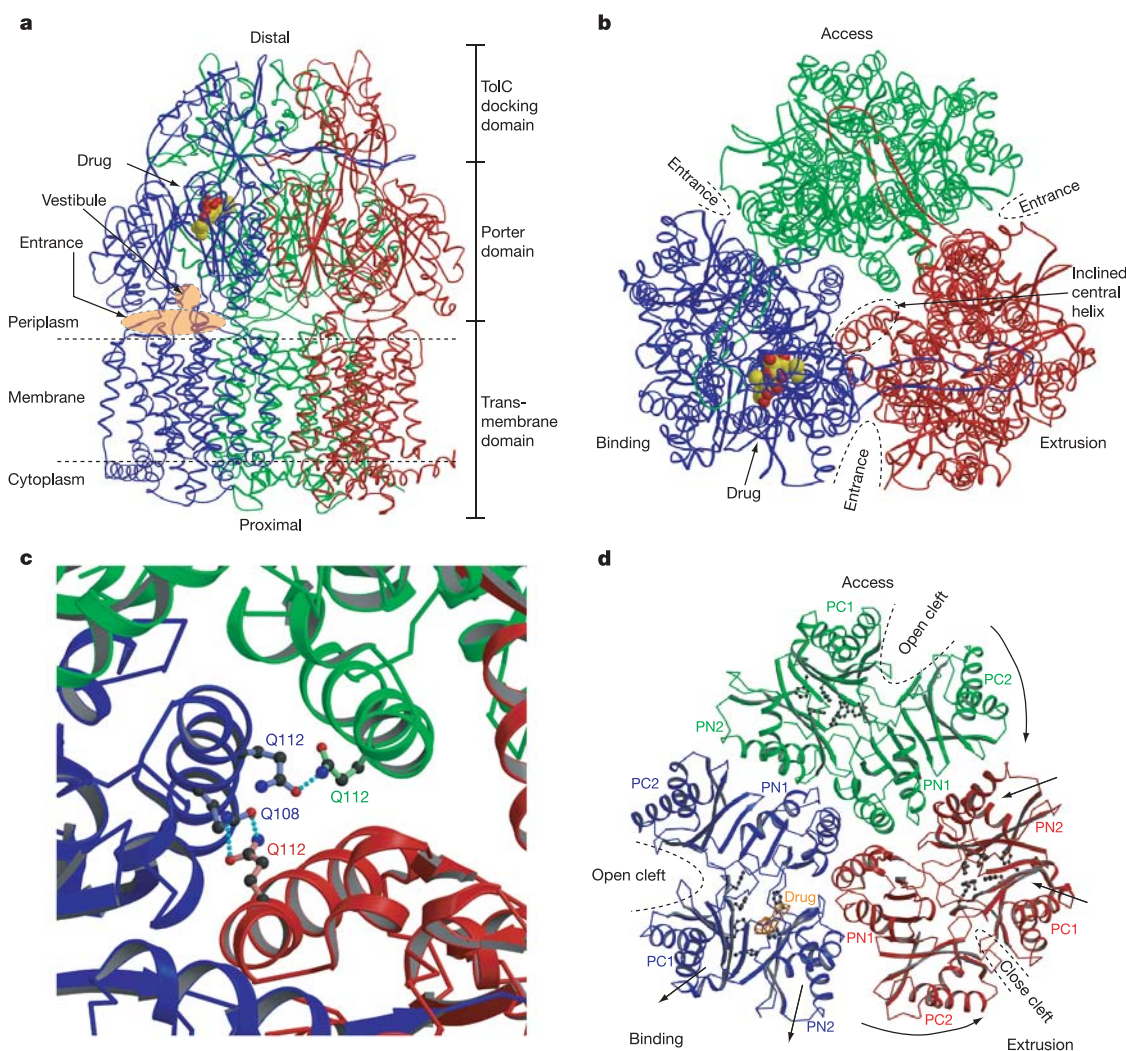


Figure 1 | Structure of the AcrB-minocycline complex based on the asymmetric crystal. The three AcrB protomers are individually coloured (blue, red and green). **a**, Ribbon representation viewed from the side parallel to the membrane plane. The minocycline molecule is shown in CPK (Corey–Pauling–Koltun) representation, with the position of the C, N and O atoms indicated by yellow, blue and red balls, respectively. The extra-membrane (periplasmic) headpiece is at the top and the transmembrane region is at the bottom. **b**, Top view of the ribbon representation. **c**, Close-up view of the

central helix in the porter domain. Key residues for the central helix inclination are shown in the ball-and-stick representation. Hydrogen bonds are indicated by dotted lines. **d**, Cut view of the porter domain from the distal side of the cell. Phenylalanine residues, including those in the substrate binding pocket, are shown in the ball-and-stick representation. The difference Fourier map ($F_{(\text{drug})} - F_{(\text{native})}$) of bound minocycline is depicted by an orange-coloured cage contour at 3σ ($0.3058 \text{ electrons } \text{\AA}^{-3}$). Arrows indicate the relative movements of the subdomains in comparison with those in the former state.

of substrates. Such multi-site binding in AcrB is similar to that seen in QacR, a soluble multidrug-binding transcriptional regulator from *Staphylococcus aureus*^{23–25}. They both have a voluminous binding pocket and primarily use aromatic residues in different positions to recognize a wide variety of hydrophobic substrates.

A conformational difference is observed between the substrate binding sites of the binding and extrusion protomers (Fig. 1d). The difference is caused by the movement of PN2 and PC1. In the binding protomer, the binding pocket is expanded owing to the movement of PN2 and PC1 away from PN1 and PC2 (Fig. 1d). The arrangement of the aromatic rings of phenylalanine residues in this expanded pocket is favourable for hydrophobic interaction with substrates (Fig. 1d and Fig. 2b, c). On the other hand, in the extrusion protomer, PN2 and PC1 are moved back towards PN1 and PC2. As a result, the space in the pocket is greatly decreased and the phenylalanine side chains are moved towards each other so that substrate binding is made impossible. Phenylalanine residues are stacked with each other presumably to stabilize the unliganded form with aromatic–aromatic interactions²⁶. Such a series of rearrangements around the binding site is also assisted by the movement of PC2 towards PC1. The vacant binding pocket of the access protomer has a shrunken structure similar to that of the extrusion protomer (Fig. 1d).

Rearrangement of the central α -helices

As shown in Fig. 1d and Fig. 2, the expanded binding pocket in the binding protomer has a gap between PN1 and PN2. This gap is a possible exit from the binding pocket. However, in the binding

protomer, the inclined helix from the extrusion protomer fills this gap and sterically blocks exit from the binding pocket (Fig. 2). At the same time, the inclination of the helix creates a space between PN1 and PN2 in the extrusion protomer, opening to the top funnel (Fig. 1d). In addition, the drug binding pocket of the extrusion protomer is shrunken so that the substrates are squeezed out. Thus, the central helices have an important role as a valve to open and shut the exit from the binding pocket.

The inclination of the central helices is mainly due to hydrogen bonding between Gln 112 in the neighbouring protomers. Gln 112 residues in the two comparably upright central helices of the access and binding protomers come close to form a hydrogen bond (Fig. 1c). In contrast, Gln 112 of the extrusion protomer forms a hydrogen bond with Gln 108 of the binding protomer. Because this Gln 108 residue is located one turn proximal to the membrane along the helix, the central helix of the extrusion protomer has to incline to match the level of the side chains (Fig. 1c).

The drug extrusion pathway

The internal solvent-accessible surface of the periplasmic region of AcrB is shown in Fig. 3a. At the proximal side of the porter domain, there are channels extending from the three entrances open to the periplasm (Fig. 1a, b). They run parallel to the membrane plane, along the interface of each protomer, and reach the central cavity. Substrates may freely access this space. At the distal end of AcrB, there is a large funnel-like opening, which connects to the TolC channel when TolC is docked at this site during drug export (Fig. 3a).

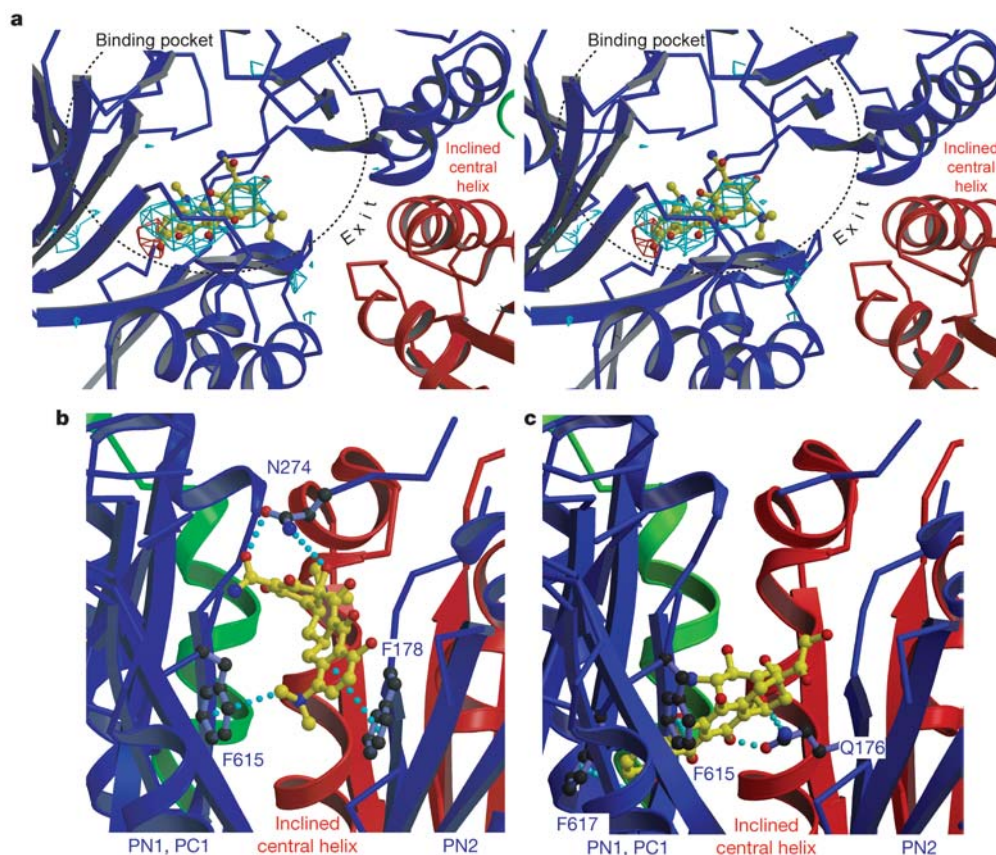


Figure 2 | Multidrug recognition by AcrB. **a**, The drug binding pocket (indicated by a dotted circle) in the porter domain with minocycline in stereo view. The difference Fourier map ($F_{(\text{drug})} - F_{(\text{native})}$) of bound minocycline at 3σ (0.3058 electrons $\text{\AA}^{-3}$; cyan), and the anomalous difference Fourier peak from 9-bromo-minocycline at 5σ (0.0195 electrons $\text{\AA}^{-3}$; red). **b**, **c**, Close-up view of the drug binding site from the cleft at the side of the

trimer. Residues 274–282, 583–600, 607–614 and 623–631 are removed for a clear view of the binding site. Drug molecules and key residues for substrate binding are shown in a ball-and-stick representation. Bound drug molecules are coloured as in Fig. 1. Chemical bonds for drug binding are drawn as dotted lines. **b**, AcrB–minocycline complex. **c**, AcrB–doxorubicin complex.

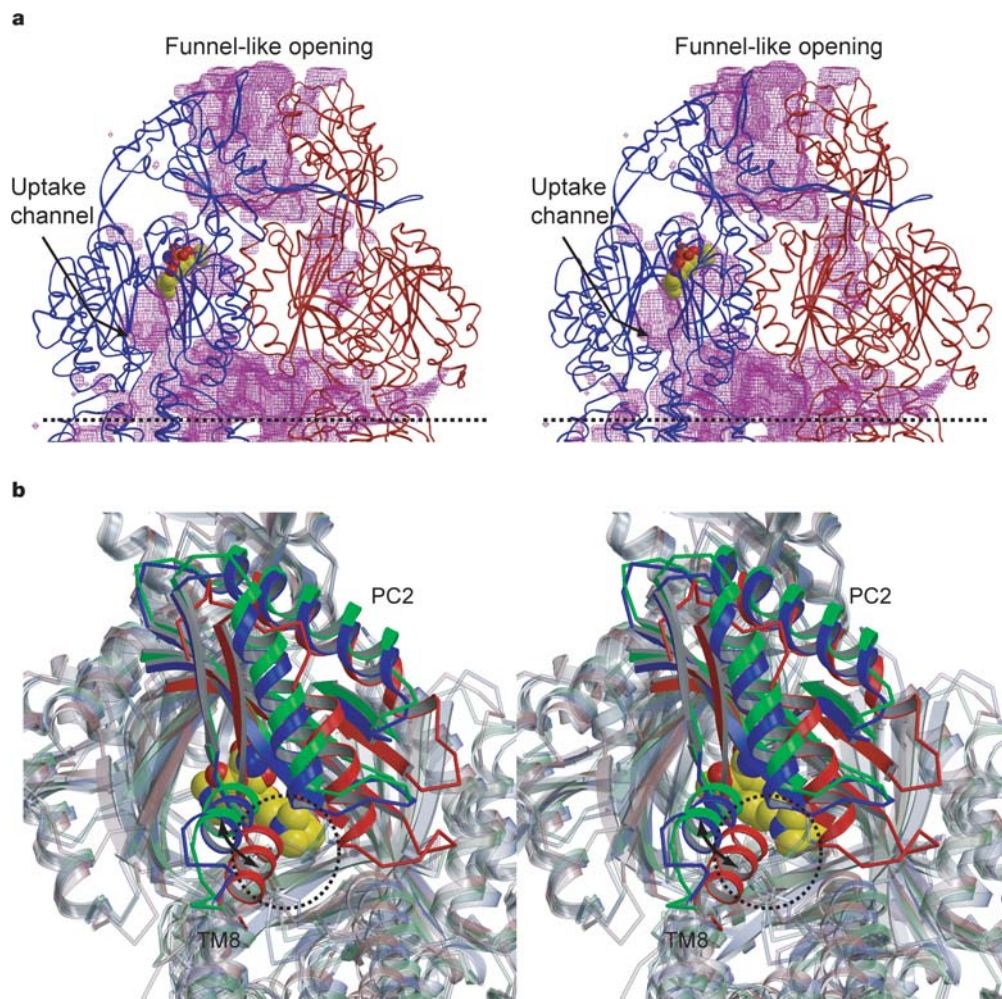


Figure 3 | The novel drug translocation pathway for AcrB. Bound minocycline in the binding protomer is shown in a CPK representation and coloured as in Fig. 1. **a**, Stereo-pair of the inner cavities inside the periplasmic region of AcrB viewed from the side parallel to the membrane plane as in Fig. 1a. The chicken-wire representation coloured in magenta is the solvent-accessible inner cavity of the AcrB molecule created with the program VOIDOO⁴² and MAMA⁴³ of the Uppsala Software Factory (<http://xray.bmc.uu.se/usf/>) with slight modifications (surface-noise

removal, and so on). For clarity, the access protomer is removed. **b**, Close-up stereo views of the vestibules in three different transport states, viewed from slightly diagonally below to the side of the molecule. Three protomers are superposed using the least-squares superposition program LSQKAB, from the CCP4 program suite³³. The flexible region above TM8 and the PC2 subdomain are represented by solid colours corresponding to the colours in Fig. 1, and the remaining main-chain tracings are more transparent. The vestibule is indicated as a dotted circle.

In the binding protomer, the substrate reaches its binding site through an uptake channel (Fig. 3a); this uptake channel is branched from the channel parallel to the membrane at a “vestibule” near the entrance, and continues through the space between the β -sheets in the porter domain to the binding pocket. Notably, in this binding protomer, the possible path from the binding pocket to the top funnel is blocked by the inclined central helix of the extrusion protomer. In contrast, in the extrusion protomer, this uptake channel has disappeared. Instead, the path from the vacant binding site to the top funnel is connected because the central α -helix is inclined away towards the binding protomer, as described above.

The opening and shutting of the pathway branched at the vestibule is caused by the movement of PC2 and partial unwinding and rewinding of the upper end of transmembrane helix 8 (TM8) (Fig. 3b). In the access protomer (Fig. 3b, green), approximately two turns of the α -helix unwind at the periplasmic end of TM8 that is connected to the β -strand in PC2, and the unwound part is moved together with PC2 (Fig. 1d). This rearrangement causes opening of the path from the vestibule. In contrast, in the extrusion protomer (Fig. 3b, red), PC2 moves towards PC1, and the consequent rewind-

ing of the α -helix at the top of TM8 shuts the vestibule leading to the uptake channel (Fig. 1d).

Transmembrane region

Drug export mediated by AcrB is driven by the proton motive force²⁷. Thus, the energization of the transport machinery must occur in the transmembrane region and transmitted to the porter domain. In the transmembrane region of each protomer of AcrB, there is only one possible proton translocation pathway: a triplet composed of Asp 407, Asp 408 and Lys 940 (Fig. 4). There are no other charged residues in the AcrB transmembrane region. These three residues are already known to be essential for AcrB function^{13,28}.

In the transmembrane region of the access and binding protomers, the Lys 940 residue is coordinated by salt bridges with Asp 407 and Asp 408 (Fig. 4). In contrast, in the extrusion protomer, Lys 940 is turned nearly 45° towards Thr 978 of TM11 and the salt bridges are abolished (Fig. 4). This movement causes the twisting of TM4 and TM10. However, how this twisting of transmembrane helices causes the movements of subdomains (including those in the porter domain), rewinding of the upper part of TM8, and the inclination

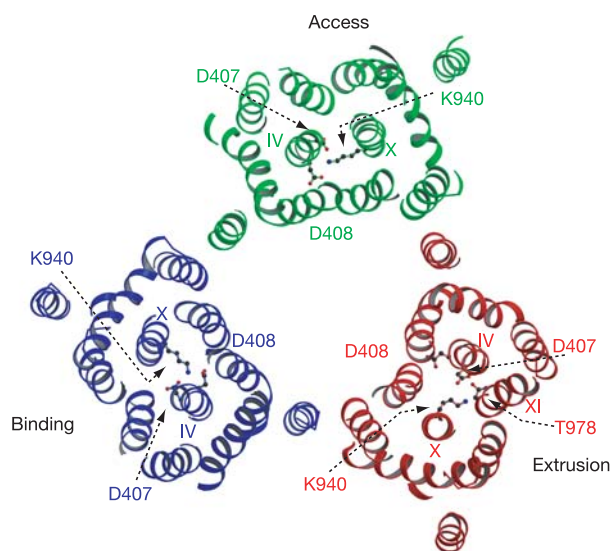


Figure 4 | Structure with a slab (~23 Å) of the transmembrane domain viewed from the periplasmic side. The side chains of three functionally essential charged residues—Asp 407, Asp 408 and Lys 940—and functionally important residue Thr 978 are shown in a ball-and-stick representation. Colours of the protomers are as in Fig. 1. Roman numerals indicate the transmembrane helix numbers¹³.

of the central α -helix is still unclear. When Thr 978 is replaced with alanine, the resulting mutant significantly loses the drug resistance phenotype (S.M., F. Ishibashi and A.Y., unpublished data).

In the crystal structure of LacY, a protonated glutamate residue in the proton translocation pathway seems to take part in substrate binding via an arginine residue that directly binds to the substrate²⁹. In contrast, the proton translocation pathway in AcrB is completely separated from the substrate translocation pathway in the periplasm.

Proposed drug export mechanism

Based on three different conformations observed for the protomers of AcrB, we propose that AcrB transports drugs by a three-step binding change mechanism (Fig. 5; see also Supplementary Table S2). In the access state, the vestibule is open to the periplasm, but the binding site is still shrunken in size. In this state, potential substrates have access to the vestibule. In the binding state, the vestibule is kept open and the binding pocket is expanded to accommodate the substrate. Therefore, drugs enter into the vestibule from the surface of the cytoplasmic membrane, move through the uptake channel, and bind to the different locations in the voluminous aromatic pocket^{23–25}. At this stage, the exit from the binding site is blocked by the central helix inclined from the extrusion protomer. Then, in the extrusion state, the vestibule is closed, and the exit is opened because the central helix is inclined away. The bound drug is pushed out into the top funnel by shrinking of the binding pocket. We expect these changes to be coupled to proton translocation across the membrane. The protonation and deprotonation of a group in the transmembrane domains would affect the accessibility or influence binding or extrusion of substrates.

The efficiency of substrate binding presumably depends on the volume and the side-chain arrangements of the binding pocket. The difference in affinity between the periplasm-open binding state and the outward-open extrusion state is the basis of the active efflux. Such an ordered binding change mechanism of the trimer is similar in principle to the ATP hydrolysis/synthesis mechanism of the $(\alpha\beta)_3$ F_1 -ATPase³⁰, except that in AcrB there is no γ -subunit that undergoes mechanical rotation. The alternate occupancy of the binding site is an essential part of the coupling mechanism. Thus, the asymmetric structure is essential for the functionality of this

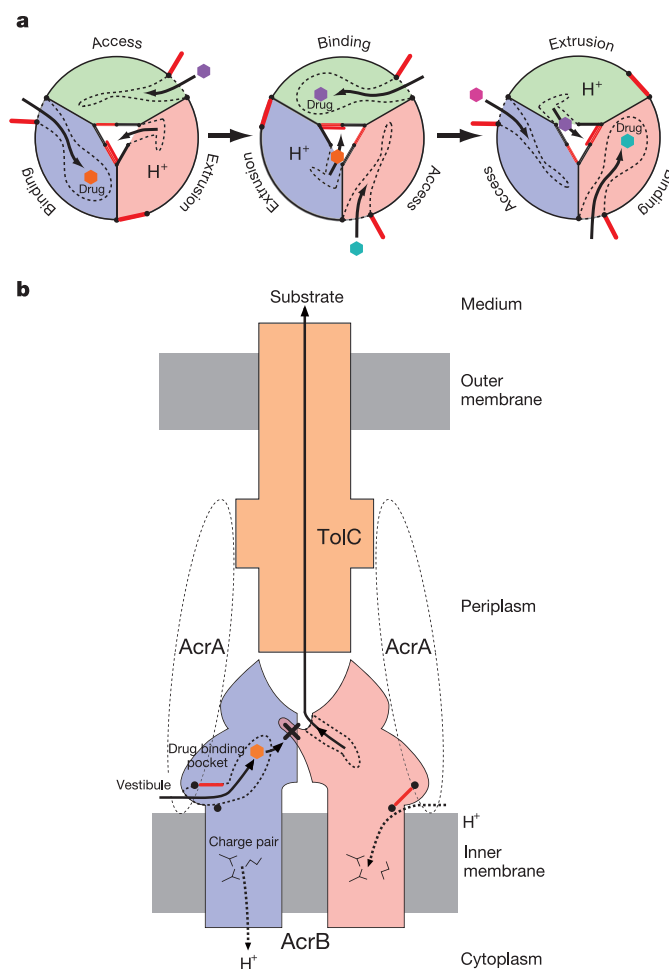


Figure 5 | Schematic illustration of the proposed 'functionally rotating ordered multidrug binding change mechanism' mediated by AcrB.

a, The top view from the distal side of the cell. Colours and arrangements in the left panel correspond to those in Fig. 1. **b**, The view from the side parallel to the membrane plane. For **a**, **b**, the entrance and exit sites within each protomer are depicted as red flaps. The drug binding pocket and translocation pathway are represented by dotted lines. Drugs are illustrated as hexagons.

transporter. This may be just a necessity of the fact that the exit pathway seems to be common to the three protomers. In the previous *R32* crystal¹³, the periplasmic region of AcrB is packed symmetrically by three neighbouring molecules around the three-fold crystallographic symmetry. Therefore, the three protomers are unavoidably fixed symmetrically by the packing force (Supplementary Fig. S3).

A conformational change is also observed around the external "cleft" between PC1 and PC2 (Fig. 1d). The cleft in the access and binding protomers is deep and wide, just as in the *R32* crystal¹³. However, the cleft has almost disappeared in the extrusion protomer, owing to the movement of PC1 and PC2 approaching each other (Fig. 1d). The cleft was suggested to be a binding site for AcrA¹³. If so, bound AcrA would have to be pushed out in the extrusion state. In other words, AcrB would push and pull the AcrA attached at the cleft during the cycle.

In contrast to the porter domain, hardly any conformational change is observed in the TolC docking domain during the binding change cycle. Thus, we speculate that the conformational change caused by proton translocation and drug binding may be transmitted through AcrA via a mutual pushing and pulling movement at the cleft. Such a rearrangement of AcrA may be transmitted to TolC to cause opening of the TolC channel.

METHODS

Preparation of native, heavy-atom derivatized and drug-complexed crystals.

Purified AcrB was prepared as described¹³. To ensure AcrB–drug complex formation, substrate (2.4 mM minocycline, 1.2 mM doxorubicin or 12 mM 9-bromo-minocycline) was added to the purified protein and incubated overnight at 4 °C (Supplementary Fig. S4). Crystals were grown by the vapour diffusion method from hanging drops at 20 °C. Protein solution containing 20 mM sodium phosphate (pH 6.0), 10% (v/v) glycerol and 0.1% (w/v) dodecanoyl sucrose were mixed (1:1) with reservoir solution containing 14% (w/v) polyethylene glycol 4000, 100 mM sodium phosphate (pH 6.2), 50 mM NaCl. The crystals were grown within one week to optimal size (0.3 × 0.3 × 0.2 mm³). The concentration of glycerol was gradually increased to 30% (v/v) by soaking in several steps for optimal cryoprotection. Crystals were picked up using nylon loops (Hampton Research) for flash-cooling in cold nitrogen gas from a cryostat (Rigaku). For heavy-atom derivatization, crystals were soaked for 1 h in cryoprotection buffer containing 0.5 mM mercuric potassium iodide, 1 mM trimethyl lead acetate, and 1 mM potassium tetrachloroplatinate. Seleno-L-methionine-labelled AcrB was prepared as described¹³ and crystallized as same as the native proteins.

Crystallographic analysis. All data sets were collected on BL41XU and BL44XU at SPring-8 with a CCD detector Quantum 315 (ADSC), and imaging-plate detector DIP6040 (Bruker AXS), respectively. All data sets were collected at 100 K. The diffraction data were processed and scaled with the HKL2000 package³¹. The native crystal belongs to the space group C2 with cell dimensions of $a = 225.9 \text{ \AA}$, $b = 134.4 \text{ \AA}$, $c = 163.2 \text{ \AA}$ and $\beta = 97.7^\circ$. Four heavy-atom derivatives of the mercury, lead, platinum and seleno-methionine-replaced crystal were isomorphous with the native crystal. Initial phases were first determined by molecular replacement using the atomic coordinates of our previous crystal structure of AcrB, which were derived from an R32 crystal form (IIWG), with the program AMoRe from the CCP4 program suite^{32,33}. The heavy-atom sites were determined in the difference Fourier maps calculated with the molecular replacement phases. Anomalous dispersion effects of the heavy metals were taken into account for the phase estimation. Multiple isomorphous replacement phases were improved and extended to 2.8 Å resolution with SHARP/autoSHARP³⁴. Model building was performed using program O³⁵, and model refinement was conducted using a semi-automatic method with program LAFIRE³⁶ with Crystallographic and NMR System (CNS)³⁷, and REFMAC³⁸. Out of 1,049 residues of the whole molecule, atomic parameters of residues 1–498 and 513–1,036 were converged well during the refinement. A crystallographic R and an R_{free} for 5% reflections excluded from the refinement were reduced to 0.268 and 0.307, respectively. To identify drug molecules in the AcrB–drug complex crystal, difference Fourier maps were calculated with coefficients of $[|F_{\text{p(drug)}}| - |F_{\text{p(native)}}|] \exp(i\alpha_{\text{p(native)}})$ at 5 Å resolution, where $F_{\text{p(drug)}}$ was the structure factor in the presence of bound substrate and $F_{\text{p(native)}}$ was that of the native crystal, and where $\alpha_{\text{p(native)}}$ was the phase of the native crystal refined at 2.8 Å resolution. A bromine atom of the brominated minocycline complex crystal was located in an anomalous difference Fourier map calculated with coefficients of $[|F_{\text{p(drug)}}^+| - |F_{\text{p(drug)}}^-|] \exp(i\alpha_{\text{p(native)}} - \pi/2)$ at 5 Å resolution, where $[|F_{\text{p(drug)}}^+| - |F_{\text{p(drug)}}^-|]$ was the Bijvoet difference of the AcrB–brominated-minocycline complex crystal measured at 0.918923 Å (bromine peak). For the two substrate complex crystals, refinement was performed at 3.1 Å resolution for the minocycline-bound form, and 3.3 Å resolution for the doxorubicin-bound form with constrained structure of these drug molecules. The quality of the diffraction data and the refinement statistics are given in Supplementary Table S1. All figures were made with MOLSCRIPT³⁹ or BOBSCRIPT⁴⁰ and rendered with RASTER3D⁴¹.

Received 10 April; accepted 11 July 2006.

Published online 16 August 2006.

- Li, X. Z. & Nikaido, H. Efflux-mediated drug resistance in bacteria. *Drugs* **64**, 159–204 (2004).
- Ma, D. *et al.* Molecular cloning and characterization of *acrA* and *acrE* genes of *Escherichia coli*. *J. Bacteriol.* **175**, 6299–6313 (1993).
- Ma, D. *et al.* Genes *acrA* and *acrB* encode a stress-induced efflux system of *Escherichia coli*. *Mol. Microbiol.* **16**, 45–55 (1995).
- Sulavik, M. C. *et al.* Antibiotic susceptibility profiles of *Escherichia coli* strains lacking multidrug efflux pump genes. *Antimicrob. Agents Chemother.* **45**, 1126–1136 (2001).
- Nishino, K. & Yamaguchi, A. Analysis of a complete library of putative drug transporter genes in *Escherichia coli*. *J. Bacteriol.* **183**, 5803–5812 (2001).
- Okusu, H., Ma, D. & Nikaido, H. AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of *Escherichia coli* multiple-antibiotic-resistant (Mar) mutants. *J. Bacteriol.* **178**, 306–308 (1996).
- Tseng, T.-T. *et al.* The RND permease superfamily: An ancient, ubiquitous and

diverse family that includes human disease and development proteins. *J. Mol. Microbiol. Biotechnol.* **1**, 107–125 (1999).

- Touze, T. *et al.* Interactions underlying assembly of the *Escherichia coli* AcrAB–TolC multidrug efflux system. *Mol. Microbiol.* **53**, 697–706 (2004).
- Poole, K. Multidrug resistance in Gram-negative bacteria. *Curr. Opin. Microbiol.* **4**, 500–508 (2001).
- Tikhonova, E. B. & Zgurskaya, H. I. AcrA, AcrB, and TolC of *Escherichia coli* form a stable intermembrane multidrug efflux complex. *J. Biol. Chem.* **279**, 32116–32124 (2004).
- Zgurskaya, H. I. & Nikaido, H. Bypassing periplasm: Reconstitution of the AcrAB multidrug efflux pump of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **96**, 7190–7195 (1999).
- Nikaido, H. How do exported proteins and antibiotics bypass the periplasm in Gram-negative bacterial cells? *Trends Microbiol.* **8**, 481–483 (2000).
- Murakami, S., Nakashima, R., Yamashita, E. & Yamaguchi, A. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419**, 587–593 (2002).
- Murakami, S. & Yamaguchi, A. Multidrug-exporting secondary transporters. *Curr. Opin. Struct. Biol.* **13**, 443–452 (2003).
- Murakami, S., Tamura, N., Saito, A., Hirata, T. & Yamaguchi, A. Extramembrane central pore of multidrug exporter AcrB in *Escherichia coli* plays an important role in drug transport. *J. Biol. Chem.* **279**, 3743–3748 (2004).
- Tamura, N., Murakami, S., Oyama, Y., Ishiguro, M. & Yamaguchi, A. Direct interaction of multidrug efflux transporter AcrB and outer membrane channel TolC detected via site-directed disulfide cross-linking. *Biochemistry* **44**, 11115–11121 (2005).
- Koronakis, V., Sharff, A., Koronakis, E., Luisi, B. & Hughes, C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405**, 914–919 (2000).
- Mikoloso, J., Bobyk, K., Zgurskaya, H. I. & Ghosh, P. Conformational flexibility in the multidrug efflux system protein AcrA. *Structure* **14**, 577–587 (2006).
- Higgins, M. K., Bokma, E., Koronakis, E., Hughes, C. & Koronakis, V. Structure of the periplasmic component of a bacterial drug efflux pump. *Proc. Natl Acad. Sci. USA* **101**, 9994–9999 (2004).
- Akama, H. *et al.* Crystal structure of the membrane fusion protein, MexA, of the multidrug transporter in *Pseudomonas aeruginosa*. *J. Biol. Chem.* **279**, 25939–25942 (2004).
- Yu, E. W., McDermott, G., Zgurskaya, H. I., Nikaido, H. & Koshland, D. E. Jr. Structural basis of multiple drug-binding capacity of the AcrB multidrug efflux pump. *Science* **300**, 976–980 (2003).
- Mao, W. *et al.* On the mechanism of substrate specificity by resistance nodulation division (RND)-type multidrug resistance pumps: The large periplasmic loops of MexD from *Pseudomonas aeruginosa* are involved in substrate recognition. *Mol. Microbiol.* **46**, 889–901 (2002).
- Schumacher, M. A. *et al.* Structural mechanisms of QacR induction and multidrug recognition. *Science* **294**, 2158–2163 (2001).
- Schumacher, M. A. & Brennan, R. G. Structural mechanisms of multidrug recognition and regulation by bacterial multidrug transcription factors. *Mol. Microbiol.* **45**, 885–893 (2002).
- Grkovic, S., Hardie, K. M., Brown, M. H. & Skurray, R. A. Interactions of the QacR multidrug-binding protein with structurally diverse ligands: Implications for the evolution of the binding pocket. *Biochemistry* **42**, 15226–15236 (2003).
- Burley, S. K. & Petsko, G. A. Aromatic–aromatic interaction: A mechanism of protein structure stabilization. *Science* **229**, 23–28 (1985).
- Thanassi, D. G., Cheng, L. W. & Nikaido, H. Active efflux of bile salts by *Escherichia coli*. *J. Bacteriol.* **179**, 2512–2518 (1997).
- Guan, L. & Nakae, T. Identification of essential charged residues in transmembrane segments of the multidrug transporter MexB of *Pseudomonas aeruginosa*. *J. Bacteriol.* **183**, 1734–1739 (2001).
- Abramson, J. *et al.* Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* **301**, 610–615 (2003).
- Boyer, P. D. A perspective of the binding change mechanism for ATP synthesis. *FASEB J.* **3**, 2164–2178 (1989).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Navaza, J. AMoRe: An automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
- Collaborative Computational Project, Number 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Yao, M., Zhou, Y. & Tanaka, I. LAFIRE: Software for automating the refinement process of protein-structure analysis. *Acta Crystallogr. D* **62**, 189–196 (2006).
- Brünger, A. T. *et al.* Crystallography & NMR System: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

38. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D* **55**, 247–255 (1999).
39. Kraulis, P. J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950 (1991).
40. Esnouf, R. M. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* **55**, 938–940 (1999).
41. Merritt, E. A. & Bacon, D. J. Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
42. Kleywegt, G. J. & Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. D* **50**, 178–185 (1994).
43. Kleywegt, G. J. & Jones, T. A. Software for handling macromolecular envelopes. *Acta Crystallogr. D* **55**, 941–944 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. J. F. Henderson, C. Toyoshima and T. Tsukihara

for discussions, advice and critical reading. We are also indebted to E. Kajitani and N. Kato for synthesis of brominated substrate. Thanks are also due to N. Shimizu, H. Sakai, M. Kawamoto, M. Yamamoto, M. Yoshimura and A. Nakagawa for data collection at SPring-8. Synchrotron experiments were performed on BL41XU and BL44XU in SPring-8 with the approval of the Japan Synchrotron Radiation Research Institute, and the Joint Research Committee of the Institute for Protein Research, Osaka University, respectively. This work was supported by PRESTO and CREST from the Japan Science and Technology Agency and grants-in-aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Author Information Coordinates for the unliganded, AcrB–minocycline complex and AcrB–doxorubicin complex structures have been deposited in the Protein Data Bank under accession numbers 2DHH, 2DRD and 2DR6, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.M. (mura@sanken.osaka-u.ac.jp).

ARTICLES

Structure of a bacterial multidrug ABC transporter

Roger J. P. Dawson¹ & Kaspar P. Locher¹

Multidrug transporters of the ABC family facilitate the export of diverse cytotoxic drugs across cell membranes. This is clinically relevant, as tumour cells may become resistant to agents used in chemotherapy. To understand the molecular basis of this process, we have determined the 3.0 Å crystal structure of a bacterial ABC transporter (Sav1866) from *Staphylococcus aureus*. The homodimeric protein consists of 12 transmembrane helices in an arrangement that is consistent with cross-linking studies and electron microscopic imaging of the human multidrug resistance protein MDR1, but critically different from that reported for the bacterial lipid flippase MsbA. The observed, outward-facing conformation reflects the ATP-bound state, with the two nucleotide-binding domains in close contact and the two transmembrane domains forming a central cavity—presumably the drug translocation pathway—that is shielded from the inner leaflet of the lipid bilayer and from the cytoplasm, but exposed to the outer leaflet and the extracellular space.

ATP-binding cassette (ABC) transporters are integral membrane proteins that actively transport chemically diverse substrates across the lipid bilayers of cellular membranes¹. Clinically relevant examples, including the human proteins MDR1 (also known as ABCB1 or P-glycoprotein) and MRP1 (also known as ABCC1), contribute to multidrug resistance of cancer cells by catalysing the extrusion of cytotoxic compounds used in cancer therapy². Other ABC proteins are associated with various diseases, including cystic fibrosis (mutations in the CFTR protein³), or have a central role in the cellular immune system by transporting antigenic peptides (TAP1 and TAP2 proteins⁴). In bacteria, ABC exporters extrude diverse substrates, including drugs and antibiotics, whereas ABC importers mediate the uptake of essential nutrients. One of the best-studied bacterial ABC exporters is the LmrA protein from *Lactobacillus lactis*, which probably contributes to bacterial drug or antibiotic resistance, but in addition has been found to functionally substitute for human MDR1 when expressed in lung fibroblast cells⁵.

The basic ABC transporter architecture consists of two transmembrane domains (TMDs) that provide a translocation pathway, and two cytoplasmic, water-exposed nucleotide-binding domains (NBDs) that hydrolyse ATP. Bacterial multidrug ABC proteins are generally expressed as 'half-transporters' that contain one TMD fused to a NBD, which dimerize to form the full transporter. Functionally important residues are highly conserved among the NBDs, suggesting that ABC transporters share a common mechanism of coupling ATP hydrolysis to substrate transport^{6,7}. However, despite extensive biochemical, genetic and structural studies, a detailed understanding of the substrate acquisition and transport process has remained elusive. To address these questions, we have determined the crystal structure of *S. aureus* Sav1866, whose primary sequence and biochemical characteristics identify it as a homologue of multidrug ABC transporters. Sav1866 shows significant sequence similarity to human ABC transporters of the subfamily B that includes MDR1 and TAP1/TAP2. Purified Sav1866 is functional in detergent solution as demonstrated by *in vitro* ATPase activity and characteristic inhibition by orthovanadate (Supplementary Fig. S1). When reconstituted in liposomes, its ATPase activity is stimulated by

the cancer drugs doxorubicin and vinblastine, and by the fluorescent dye Hoechst 33342 (Supplementary Fig. S1), all of which are also substrates of *L. lactis* LmrA and human MDR1^{8,9}.

Sav1866 architecture and domain structure

Our crystallographic analysis yielded electron density of high quality for the entire protein (Supplementary Fig. S2), and the structure was refined to 3.0 Å resolution with good refinement statistics (R_{work} and R_{free} values of 25.5% and 27.2%, respectively). The functional unit of Sav1866 is a dimer of two elongated subunits related by two-fold molecular and non-crystallographic symmetry (Fig. 1a, b). The full transporter is 120 Å long, 65 Å wide and 55 Å deep, and its overall shape is consistent with that of human MDR1 studied at low resolution by electron microscopy^{10,11}. Each subunit consists of an amino-terminal TMD (amino-acid residues 1–320) and a carboxy-terminal NBD (residues 337–578). The two subunits exhibit a considerable twist and embrace each other, with both the transmembrane and the nucleotide-binding domains tightly interacting. The polypeptide that links the domains (residues 321–336) emanates from transmembrane (TM) helix TM6 and wraps around the distal side of the NBD (Fig. 1b), similar to the analogous linker observed in the crystal structure of human TAP1¹².

The NBDs of the Sav1866 homodimer are similar in structure to those of other ABC transporters. They expose conserved ATP-binding and -hydrolysis motifs⁷ at the shared interface in a 'head-to-tail' arrangement that has been widely accepted as physiologically relevant on the basis of structural^{13,14} and biochemical^{15–17} data. The hallmark of this arrangement is that two ATP-hydrolysis sites are formed at the shared interface, between the P-loop of one NBD and the ABC signature motif of the other. This creates a direct link between the two sites and provides the molecular basis of the observed cooperativity in ATP binding and hydrolysis¹⁸.

Sav1866 was crystallized in the presence of ADP, and clear electron density was evident for two ADP molecules bound by the P-loops and the ABC signature motifs of opposing subunits. However, the conformation of these motifs is, at the present resolution, indistinguishable from that of isolated nucleotide-binding domains

¹Institute of Molecular Biology and Biophysics, ETH Zurich, 8093 Zurich, Switzerland.

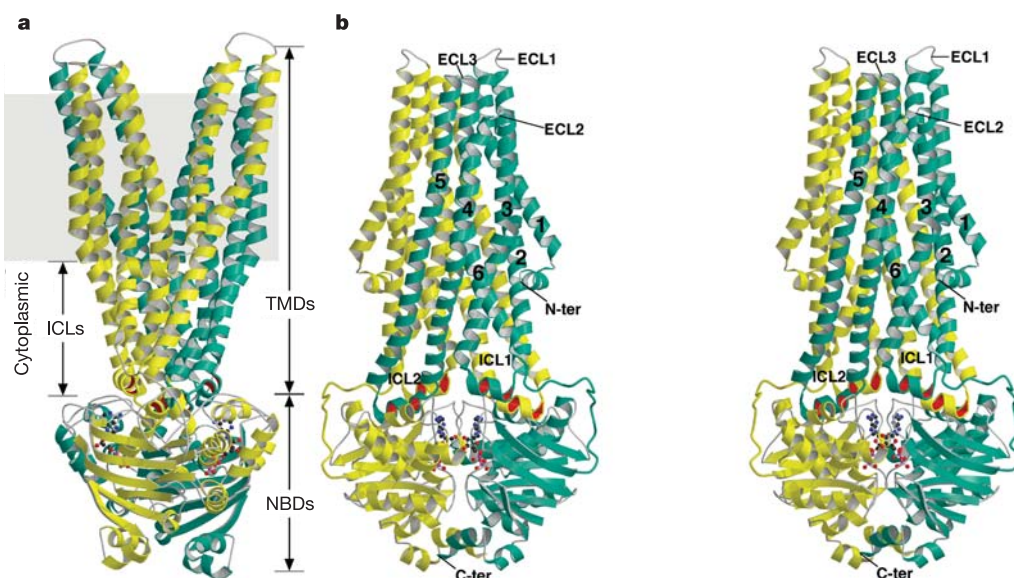


Figure 1 | Sav1866 structure. **a**, Backbone of the homodimeric protein in ribbon representation, with the subunits coloured yellow and turquoise. Bound ADP is in ball-and-stick representation. The view reveals the membrane-embedded ‘wings’ of the protein; the grey box depicts the probable location of the lipid bilayer on the basis of the hydrophobicity of the protein surface. **b**, Stereo view rotated with respect to **a** by 90° around the

vertical two-fold molecular and non-crystallographic axis. The transmembrane helices of one subunit (turquoise) are numbered. TMDs, transmembrane domains; NBDs nucleotide-binding domains; ICL, intracellular loops (between transmembrane helices); ECL, extracellular loops (between transmembrane helices); N-ter, amino terminus; C-ter, carboxy terminus.

containing trapped ATP (for example, the archaeal ABC protein MJ0796¹⁴). Specifically, the P-loops and ABC signature motifs tightly sandwich bound nucleotide between them, whereas the structure of the *Escherichia coli* vitamin B₁₂ transporter BtuCD, crystallized in the absence of nucleotide¹⁹, revealed a substantial separation of the analogous motifs (Fig. 2). This indicates that although ADP, rather than ATP, was bound, the NBDs in Sav1866 exhibit the conformation of the ATP-bound state. This is consistent with the TMDs adopting an outward-facing conformation (see below). We conclude that our purification and crystallization conditions, in particular the presence of detergent, have shifted the conformational equilibrium of Sav1866 to the ATP-bound state.

Each TMD traverses the lipid bilayer six times to yield 12 transmembrane helices for the homodimeric transporter, in agree-

ment with the canonical ABC exporter topology. Around the middle of the membrane, bundles of transmembrane helices diverge into two discrete ‘wings’ that point away from one another towards the cell exterior, thus providing an outward-facing conformation. Rather than representing individual TMDs, each ‘wing’ consists of helices TM1–TM2 from one subunit and TM3–TM6 from the other subunit. The transmembrane segments are connected by long intracellular and short extracellular loops (ICLs and ECLs, respectively). The ICLs extend the helical secondary structure beyond the lipid bilayer and protrude approximately 25 Å into the cytoplasm. Notably, the helices TM1–TM3 are closely related to TM4–TM6 by an approximate two-fold rotation around a symmetry axis parallel to the membrane plane, with a root-mean-square deviation of 5.7 Å for 146 C α positions (residues 9–154 and 175–320). Rotational symmetry of subsets of

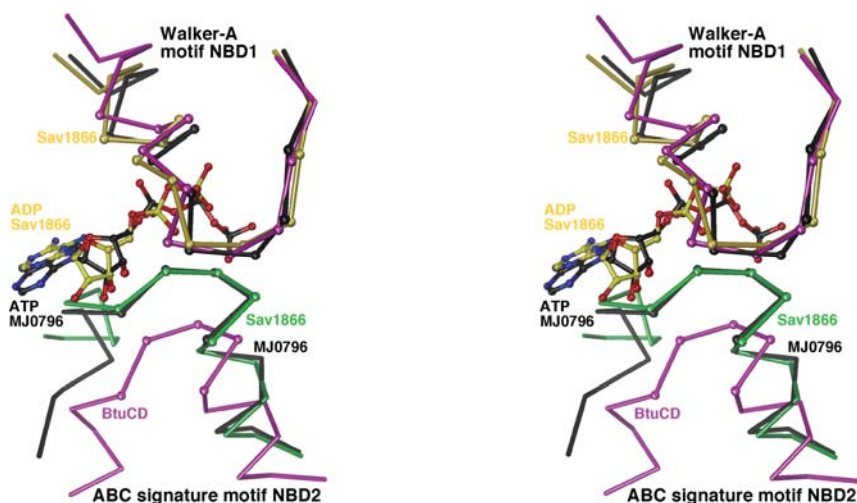


Figure 2 | Superposition of nucleotide-binding domain structures. Both P-loops (Walker-A motifs) of the NBD dimers were used for superposition, but for clarity only one nucleotide-binding site is shown in stereo. The C α positions of conserved residues GxxGxGKST (P-loop) and LSGGQ (ABC

signature motifs) are depicted as spheres in the backbone traces. Bound nucleotides are in ball-and-stick representation. Note the close superposition of the ABC signature motifs of Sav1866 and MJ0796 compared with that of BtuCD.

transmembrane helices may indicate gene duplication and has been observed in the structures of aquaporins²⁰ and major facilitators²¹, but has not previously been reported for ABC exporters.

The arrangement of transmembrane helices observed for Sav1866 is consistent with recent cross-linking data that identified neighbouring helices in human MDR1²². In contrast, the reported X-ray analyses of another ABC exporter, the bacterial lipid flippase MsbA^{23–25}, show an arrangement distinctly different from the one we describe here (Supplementary Fig. S4 and Supplementary Table S2). With respect to the nucleotide-binding domains, the structure of Sav1866 exhibits close structural agreement with that of MJ0796¹⁴,

whereas in MsbA of *Salmonella typhimurium*, the analogous domains appear laterally offset and at a greater distance. The observed architectures of MsbA and Sav1866 remain incompatible, even when considering that the proteins may have been trapped in distinct states, and the differences—if real—would indicate a convergent evolution of the two proteins.

Transmission interface

Conformational changes generated by ATP binding and hydrolysis are transmitted from the nucleotide-binding to the transmembrane domains through non-covalent interactions at the shared interface. The TMDs of Sav1866 contribute to this interface mainly through the intracellular loops ICL1 and ICL2. This is consistent with genetic data that implicated ICL1 of human MDR1 to interact with the NBDs²⁶, and with mutational studies and sequence comparisons that indicated ICL2 of TAP1/TAP2, as well as ICL4 of CFTR, to provide similar crucial contacts^{4,27}. In Sav1866, both ICLs contain short helices oriented roughly parallel to the membrane plane and

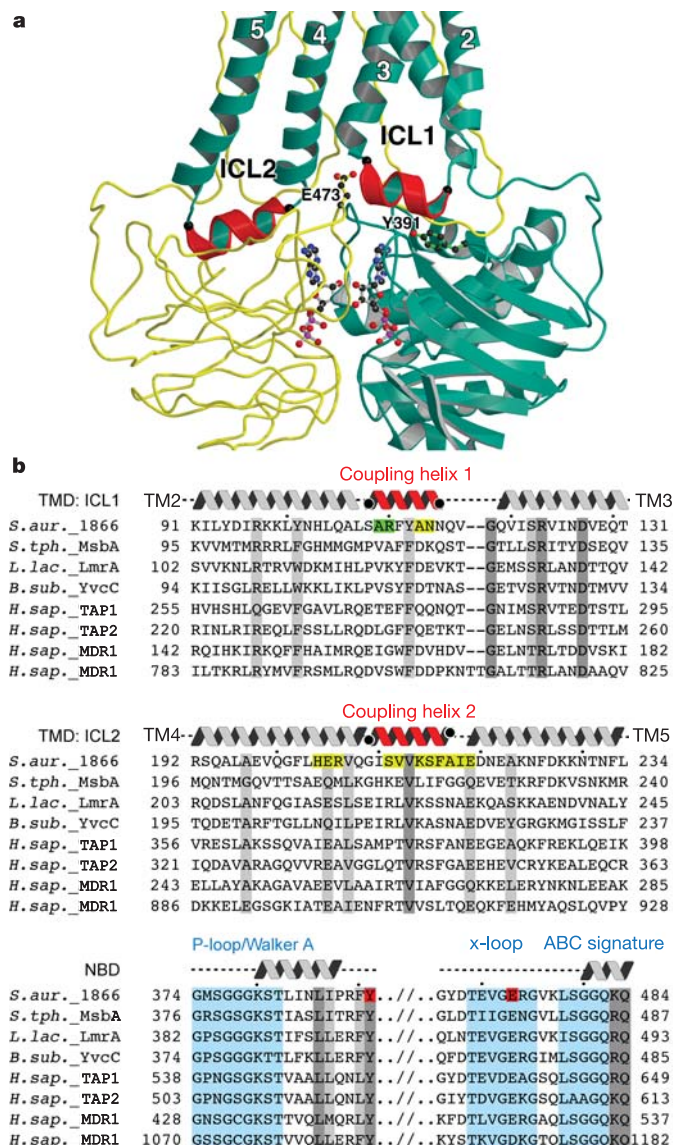


Figure 3 | Transmission interface. **a**, Close-up view of Sav1866 with one subunit in turquoise ribbon representation, the other in yellow coil representation. For clarity, only the ICL regions of the transmembrane domains are shown. Black spheres depict the first and last C α positions of the 'coupling helices' (see the text). Side chains of the conserved Glu 473 and Tyr 391 are in ball-and-stick representation. **b**, Alignment of the Sav1866 protein sequence with those of other ABC transporters. Residues with high conservation are shaded dark grey, those with significant conservation are shaded light grey, and relevant motifs are shaded blue and indicated. Residues of the ICLs interacting (4 Å cut-off) with the NBD of the same subunit are indicated with green shading, and those contacting the opposite NBD with yellow shading. Residues of the NBD interacting with the TMD are shown with red shading.

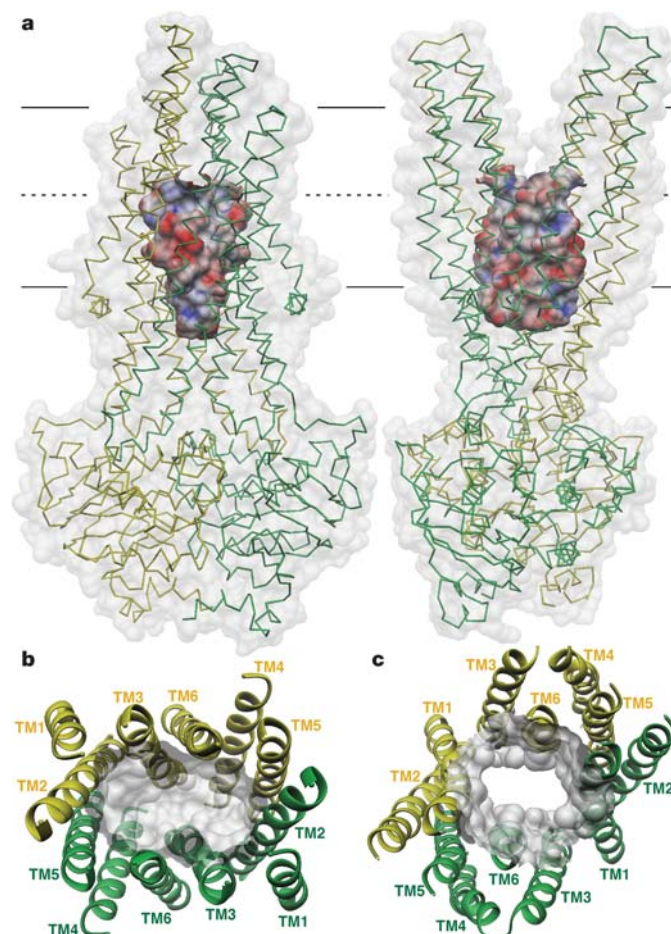


Figure 4 | Substrate translocation pathway. **a**, Backbone traces of the Sav1866 subunits (green and yellow) in two orientations, rotated by 90° around a vertical axis. The molecular surface of the central cavity is coloured according to its calculated electrostatic potential (blue, positively charged; red, negatively charged), whereas that of the rest of the transporter is in transparent light grey. Lines indicate the approximate position of the two leaflets of the lipid bilayer. Access from the cavity to the outer leaflet of the lipid bilayer is visible in the right panel from the front and back, between the 'wings' formed by the TMDs. **b**, Cavity at the level of the inner leaflet viewed from the extracellular side. The transmembrane helices are numbered and the cavity is shown as a grey surface. **c**, Same as **b** but at the level of the outer leaflet. Note that owing to helix bending, different subsets of helices line the cavity when compared with **b**.

providing the bulk of the contacts (Fig. 3a,b). We refer to these as 'coupling helices' to emphasize their likely involvement in transmitting mechanistically crucial conformational changes. Unexpectedly, the ICLs of Sav1866 reach across and contact primarily the nucleotide-binding domain of the opposite subunit, analogous to domain swapping observed in various enzymes²⁸. Although coupling helix 1 contacts the NBDs of both subunits, coupling helix 2 interacts with that of the opposite subunit exclusively. Such swapping has not been anticipated for ABC exporters (Fig. 5), but is not inconsistent with the available biochemical or genetic data. In contrast, the TMDs of the ABC importer BtuCD contact only one NBD, resulting in a large gap at the centre of the four protein domains¹⁹. The architecture of the ICLs and coupling helices observed in Sav1866 is probably conserved among ABC exporters, as suggested by the significant sequence similarity in the corresponding regions (Fig. 3b).

On the part of the NBDs, the contact surface with the TMD is primarily lined with residues around the so-called 'Q-loop'⁷, as was also observed in BtuCD¹⁹. Two prominent exceptions include the conserved residues Tyr 391 and Glu 473. The latter seems particularly intriguing, as it interacts with both ICLs and is part of a previously unrecognized, short sequence motif (TEVGERG) that appears conserved in ABC export proteins only (Fig. 3b). We refer to this sequence as the 'x-loop' to highlight its apparent function in cross-linking the ICLs. Because the x-loops precede the ABC signature motifs, they probably respond to ATP binding and hydrolysis and may transmit conformational changes to the ICLs by alternating engaging in, and releasing, the cross-link on ATP binding and hydrolysis. The absence of this motif in binding-protein-dependent bacterial ABC importers suggests a distinct coupling mechanism compared with ABC exporters.

Substrate translocation pathway

A large cavity is present at the interface of the two transmembrane domains (Fig. 4). Although shielded from the cytoplasm and the lipid bilayer at the level of the inner leaflet, the cavity is accessible from the outer leaflet and exposed to the extracellular space. The bottom of the cavity reaches beyond the intracellular membrane boundary, but no connection to the cytoplasm exists. At the level of the inner leaflet, the cavity features a hydrophilic surface that is primarily lined with polar and charged amino-acid side chains from helices TM2–TM5, with a slight surplus of negative charges and no significant hydrophobic patches. At the level of the outer leaflet, the translocation pathway is lined with residues from helices TM1, TM3 and TM6.

The observed cavity is consistent with an outward-facing conformation of the ATP-bound state of human MDR1, as revealed by electron microscopy^{10,29}. The predominance of polar and charged

amino acids suggests that rather than a high-affinity binding site, the observed cavity in Sav1866 may reflect an extrusion pathway with little or no affinity for hydrophobic drugs. Binding of substrates to human MDR1 was indeed significantly reduced in the presence of ATP³⁰. Similarly, LmrA exposes a high-affinity binding site to the interior of the cell, whereas in the ATP-bound state, this site is occluded and a weak binding site is accessible from the extracellular space⁹.

The architecture of Sav1866 indicates that residues from all transmembrane helices contribute to the surface of the translocation pathway. It was noted recently that, despite extensive efforts, no well-defined substrate binding sites have been identified in ABC exporters³¹. Residues from transmembrane helices TM1, TM4–TM6 and TM10–TM12 have been implicated in drug binding to human MDR1³², whereas residues from multiple helices have been found to contribute to substrate binding to TAP1/TAP2³³. The analogous residues in Sav1866 indeed point towards the translocation pathway.

Transport mechanism

The simplest scheme of the transport mechanism invokes two states: an inward-facing conformation with the substrate binding site accessible from the cell interior, and an outward-facing conformation with an extrusion pocket exposed to the external medium. The structure of Sav1866 reveals that tight interaction of the NBDs in the ATP-bound state is coupled to the outward-facing conformation of the TMDs. In this conformation, bound substrates may escape into the outer leaflet of the lipid bilayer or into the aqueous medium surrounding the cell, depending on their hydrophobicity. Hydrolysis of ATP is expected to return the transporter to an inward-facing conformation, again granting access to the binding site from the cell interior. ABC transporters may thus use an "alternating access and release" mechanism first postulated for major facilitator transport proteins³⁴, with the distinction that ATP binding and hydrolysis, rather than substrate acquisition, may control the conversion of one state into the other. Most ABC transporters bind and hydrolyse two ATP molecules in each reaction cycle^{35,36}. However, the number of bound substrates can vary, and—depending on the molecular mass—up to two substrates have been found to enter the binding pockets of human MDR1 and MRP2^{37,38}. We conclude that if two small substrates are bound, the apparent stoichiometry of hydrolysed ATP per transported substrate will be one, whereas for a single, larger substrate, it will be two.

The architecture of the TMD–NBD interface of Sav1866 defies the widely used ABC exporter schematic shown in Fig. 5a. Rather than aligned side-by-side, the two subunits of Sav1866 are intricately associated (Fig. 5b) and the main interface of the intracellular loops is with the NBD of the opposite subunit. Furthermore, helices from both TMDs form the two 'wings' consisting of bundles of helices embedded in the membrane (Fig. 1a). Given these constraints, the two subunits are unlikely to move independently and their maximum separation during the reaction cycle is therefore limited. Our results challenge mechanistic models that suggest that NBDs dimerize on binding ATP and dissociate on completing a transport cycle^{31,39,40}. Although this is true for purified, isolated NBDs, electron microscopic studies of full-length MDR1 and CFTR have demonstrated close proximity of the NBDs in the absence of nucleotide^{11,41}, as has the crystallographic analysis of BtuCD¹⁹.

Compared with Sav1866, the ABC signature motifs of the nucleotide-free state of BtuCD are separated by an additional ~5 Å from the P-loops (Fig. 2), a conformational rearrangement that reduces the interface between opposing NBDs. Solvent accessibility of the nucleotide-binding site in the bacterial maltose transporter is indeed decreased in the absence of ATP⁴², as is the energetic coupling between side chains at the NBD interface of CFTR¹⁷. The 'power stroke' of ABC transporters, triggered by binding of ATP, may thus consist of moderate conformational rearrangements that originate at the interface of the two NBDs and propagate towards the TMDs.

In addition to its value for investigating the drug extrusion

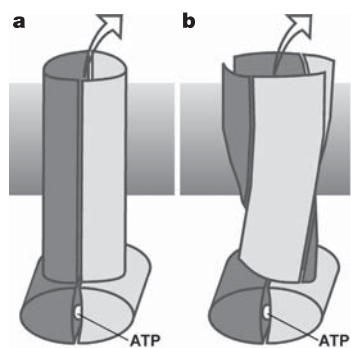


Figure 5 | ABC exporter schematics. **a**, Earlier cartoons depict two compact transporter halves (subunits) arranged side-by-side, suggesting separation during the transport cycle. The grey box indicates the location of the membrane. **b**, Schematic of Sav1866 in the observed, outward-facing conformation. The cartoon emphasizes the domain swapping and subunit twisting. Arrows indicate the release of bound drug into the extracellular space.

mechanism, the Sav1866 structure may be useful for interpreting mutations of human ABC transporters of known substrates or proven implication in disease. A notable example is the deletion-mutation $\Delta 508$ (a Phe residue) of CFTR, responsible for 70% of cystic fibrosis cases^{43,44}. The analogous residue in Sav1866 is revealed as Leu 437 by superposition with the NBD of wild-type and $\Delta 508$ -mutant CFTR⁴⁵. Leu 437 is part of a cluster of aromatic and hydrophobic side chains at the interface of ICL2 and the NBD. Given this structural context, the absence of the Phe side chain in CFTR $\Delta 508$, despite limited impact on the protein backbone, probably destabilizes the analogous packing of residues at the interface of NBD1 and ICL4, thus contributing to the observed misassembly of the mutant CFTR protein.

Conclusions

Our results define, at high resolution, the architecture of a bacterial multidrug ABC transporter in the outward-facing conformation, with a single substrate translocation pathway exposed to the extracellular environment. Because the structure of Sav1866 agrees well with genetic, biochemical and structural data of bacterial and human homologues, it most likely reflects a physiologically relevant state. It visualizes the arrangement of the canonical 12 transmembrane helices forming the core of ABC exporter proteins. Given that high-resolution structures of clinically relevant ABC transporters are currently not available, Sav1866 may serve as a valuable structural model of human homologues and may initiate the rational design of drugs aimed at interfering with the extrusion of agents used in chemotherapy. The challenge for future studies will be to trap and visualize the transporter in the inward-facing conformation that allows substrates to access the high-affinity binding site.

METHODS

Sav1866 was selected from a pool of bacterial ABC exporter homologues on the basis of its superior biochemical stability and its ability to form well-ordered, three-dimensional crystals. The gene encoding Sav1866 was cloned from *S. aureus* genomic DNA by polymerase chain reaction (PCR) and fused to an N-terminal decahistidine affinity tag in a modified T7 expression vector. The transporter was overexpressed in *E. coli*, solubilized from isolated cell membranes, and purified in octaethylene glycol monododecyl ether ($C_{12}E_8$) as a detergent. Protein crystals were obtained by vapour diffusion and belonged to the space group C2 with one full transporter per asymmetric unit. Diffraction data were collected at the beamline X06SA of the Swiss Light Source (SLS). The structure was phased by multiple isomorphous replacement with anomalous scattering using data from various derivative crystals (Supplementary Table S1). Excellent experimental density was visible for the entire molecule, including loops connecting transmembrane helices (Supplementary Fig. S2), and model building was further aided by the presence of selenomethionines serving as marker positions (Supplementary Fig. S3). Except for extracellular loops engaged in lattice contacts, and a few residues with distinct side-chain conformations at the subunit interface, no differences are visible between the two non-crystallographically related subunits. Details of the biochemical and crystallographic procedures are given in Supplementary Information. Figures 1 and 2 were prepared using MolScript (<http://www.avatar.se/molscript>) and Raster3D (<http://skuld.bmsc.washington.edu/raster3d>), Figs 3 and 4 using DINO (<http://www.dino3d.org>) and MSMS (http://www.scripps.edu/~sanner/html/msms_home.html). Electrostatic potentials were calculated using MEAD (<http://www.scripps.edu/mb/bashford>).

Received 9 May; accepted 11 August 2006.

Published online 30 August 2006.

- Holland, I. B., Cole, S. P. C., Kuchler, K. & Higgins, C. F. *ABC Proteins: From Bacteria to Man* (Academic, London, 2003).
- Gottesman, M. M. & Ambudkar, S. V. Overview: ABC transporters and human disease. *J. Bioenerg. Biomembr.* **33**, 453–458 (2001).
- Sheppard, D. N. & Welsh, M. J. Structure and function of the CFTR chloride channel. *Physiol. Rev.* **79**, S23–S45 (1999).
- Lankat-Buttgereit, B. & Tampe, R. The transporter associated with antigen processing: function and implications in human diseases. *Physiol. Rev.* **82**, 187–204 (2002).
- van Veen, H. W. *et al.* A bacterial antibiotic-resistance gene that complements the human multidrug-resistance P-glycoprotein gene. *Nature* **391**, 291–295 (1998).
- Holland, I. B. & Blight, M. A. ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. *J. Mol. Biol.* **293**, 381–399 (1999).
- Schneider, E. & Hunke, S. ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains. *FEMS Microbiol. Rev.* **22**, 1–20 (1998).
- Shapiro, A. B. & Ling, V. Reconstitution of drug transport by purified P-glycoprotein. *J. Biol. Chem.* **270**, 16167–16175 (1995).
- van Veen, H. W., Margolles, A., Muller, M., Higgins, C. F. & Konings, W. N. The homodimeric ATP-binding cassette transporter LmrA mediates multidrug transport by an alternating two-site (two-cylinder engine) mechanism. *EMBO J.* **19**, 2503–2514 (2000).
- Rosenberg, M. F., Callaghan, R., Modok, S., Higgins, C. F. & Ford, R. C. Three-dimensional structure of P-glycoprotein—the transmembrane regions adopt an asymmetric configuration in the nucleotide-bound state. *J. Biol. Chem.* **280**, 2857–2862 (2005).
- Lee, J. Y., Urbatsch, I. L., Senior, A. E. & Wilkens, S. Projection structure of P-glycoprotein by electron microscopy—evidence for a closed conformation of the nucleotide binding domains. *J. Biol. Chem.* **277**, 40125–40131 (2002).
- Gaudet, R. & Wiley, D. C. Structure of the ABC ATPase domain of human TAP1, the transporter associated with antigen processing. *EMBO J.* **20**, 4964–4972 (2001).
- Hopfner, K. P. *et al.* Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* **101**, 789–800 (2000).
- Smith, P. C. *et al.* ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol. Cell* **10**, 139–149 (2002).
- Hrycyna, C. A. *et al.* Mechanism of action of human P-glycoprotein ATPase activity—photochemical cleavage during a catalytic transition state using orthovanadate reveals cross-talk between the two ATP sites. *J. Biol. Chem.* **273**, 16631–16634 (1998).
- Chen, J., Sharma, S., Quirocho, F. A. & Davidson, A. L. Trapping the transition state of an ATP-binding cassette transporter: evidence for a concerted mechanism of maltose transport. *Proc. Natl Acad. Sci. USA* **98**, 1525–1530 (2001).
- Vergani, P., Lockless, S. W., Nairn, A. C. & Gadsby, D. C. CFTR channel opening by ATP-driven tight dimerization of its nucleotide-binding domains. *Nature* **433**, 876–880 (2005).
- Senior, A. E. & Bhagat, S. P-glycoprotein shows strong catalytic cooperativity between the two nucleotide sites. *Biochemistry* **37**, 831–836 (1998).
- Locher, K. P., Lee, A. T. & Rees, D. C. The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* **296**, 1091–1098 (2002).
- Stroud, R. M. *et al.* Glycerol facilitator GlpF and the associated aquaporin family of channels. *Curr. Opin. Struct. Biol.* **13**, 424–431 (2003).
- Abramson, J., Kaback, H. R. & Iwata, S. Structural comparison of lactose permease and the glycerol-3-phosphate antiporter: members of the major facilitator superfamily. *Curr. Opin. Struct. Biol.* **14**, 413–419 (2004).
- Stenham, D. R. *et al.* An atomic detail model for the human ATP binding cassette transporter P-glycoprotein derived from disulphide cross-linking and homology modeling. *FASEB J.* **17**, 2287–2289 (2003).
- Chang, G. & Roth, C. B. Structure of MsbA from *E. coli*: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science* **293**, 1793–1800 (2001).
- Chang, G. Structure of MsbA from *Vibrio cholera*: a multidrug resistance ABC transporter homolog in a closed conformation. *J. Mol. Biol.* **330**, 419–430 (2003).
- Reyes, C. L. & Chang, G. Structure of the ABC transporter MsbA in complex with ADP-vanadate and lipopolysaccharide. *Science* **308**, 1028–1031 (2005).
- Currier, S. J. *et al.* Identification of residues in the first cytoplasmic loop of P-glycoprotein involved in the function of chimeric human MDR1-MDR2 transporters. *J. Biol. Chem.* **267**, 25153–25159 (1992).
- Cotten, J. F., Ostedgaard, L. S., Carson, M. R. & Welsh, M. J. Effect of cystic fibrosis-associated mutations in the fourth intracellular loop of cystic fibrosis transmembrane conductance regulator. *J. Biol. Chem.* **271**, 21279–21284 (1996).
- Liu, Y. & Eisenberg, D. 3D domain swapping: as domains continue to swap. *Protein Sci.* **11**, 1285–1299 (2002).
- Rosenberg, M. F. *et al.* Repacking of the transmembrane domains of P-glycoprotein during the transport ATPase cycle. *EMBO J.* **20**, 5615–5625 (2001).
- Ramachandra, M. *et al.* Human P-glycoprotein exhibits reduced affinity for substrates during a catalytic transition state. *Biochemistry* **37**, 5010–5019 (1998).
- Higgins, C. F. & Linton, K. J. The ATP switch model for ABC transporters. *Nature Struct. Mol. Biol.* **11**, 918–926 (2004).
- Loo, T. W. & Clarke, D. M. Recent progress in understanding the mechanism of P-glycoprotein-mediated drug efflux. *J. Membr. Biol.* **206**, 173–185 (2005).
- Nijenhuis, M. & Hammerling, G. J. Multiple regions of the transporter associated with antigen processing (TAP) contribute to its peptide binding site. *J. Immunol.* **157**, 5467–5477 (1996).
- Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969–970 (1966).

35. Patzlaff, J. S., van der Heide, T. & Poolman, B. The ATP/substrate stoichiometry of the ATP-binding cassette (ABC) transporter OpuA. *J. Biol. Chem.* **278**, 29546–29551 (2003).
36. Sauna, Z. E. & Ambudkar, S. V. Evidence for a requirement for ATP hydrolysis at two distinct steps during a single turnover of the catalytic cycle of human P-glycoprotein. *Proc. Natl Acad. Sci. USA* **97**, 2515–2520 (2000).
37. Shapiro, A. B. & Ling, V. Positively cooperative sites for drug transport by P-glycoprotein with distinct drug specificities. *Eur. J. Biochem.* **250**, 130–137 (1997).
38. Zelcer, N. *et al.* Evidence for two interacting ligand binding sites in human multidrug resistance protein 2 (ATP binding cassette C2). *J. Biol. Chem.* **278**, 23538–23544 (2003).
39. Moody, J. E., Millen, L., Binns, D., Hunt, J. F. & Thomas, P. J. Cooperative, ATP-dependent association of the nucleotide binding cassettes during the catalytic cycle of ATP-binding cassette transporters. *J. Biol. Chem.* **277**, 21111–21114 (2002).
40. Chen, J., Lu, G., Lin, J., Davidson, A. L. & Quirocho, F. A. A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol. Cell* **12**, 651–661 (2003).
41. Awayn, N. H. *et al.* Crystallographic and single-particle analyses of native and nucleotide-bound forms of the cystic fibrosis transmembrane conductance regulator (CFTR) protein. *Biochem. Soc. Trans.* **33**, 996–999 (2005).
42. Mannering, D. E., Sharma, S. & Davidson, A. L. Demonstration of conformational changes associated with activation of the maltose transport complex. *J. Biol. Chem.* **276**, 12362–12368 (2001).
43. Kerem, B. S. *et al.* Identification of mutations in regions corresponding to the two putative nucleotide (ATP)-binding folds of the cystic-fibrosis gene. *Proc. Natl Acad. Sci. USA* **87**, 8447–8451 (1990).
44. Tsui, L. C. The spectrum of cystic-fibrosis mutations. *Trends Genet.* **8**, 392–398 (1992).
45. Lewis, H. A. *et al.* Impact of the $\Delta F508$ mutation in first nucleotide-binding domain of human cystic fibrosis transmembrane conductance regulator on domain folding and structure. *J. Biol. Chem.* **280**, 1346–1353 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Schulze-Briese, E. Pohl and T. Tomizaki for assistance with synchrotron data collection, D. Sargent for help with xenon derivatization, and J. Rosenbusch for critical reading of the manuscript. This work was supported by the Roche Research Fund, the NCCR Structural Biology Zurich, the Swiss National Science Foundation, and the Swiss Cancer League Oncosuisse.

Author Information Coordinates and structure factors for Sav1866 have been deposited in the Protein Data Bank under the accession code 2HYD. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.P.L. (kaspar.locher@mol.biol.ethz.ch).

LETTERS

A galaxy at a redshift $z = 6.96$

Masanori Iye^{1,2,3}, Kazuaki Ota², Nobunari Kashikawa¹, Hisanori Furusawa⁴, Tetsuya Hashimoto², Takashi Hattori⁴, Yuichi Matsuda⁵, Tomoki Morokuma⁶, Masami Ouchi⁷ & Kazuhiro Shimasaku²

When galaxy formation started in the history of the Universe remains unclear. Studies of the cosmic microwave background indicate that the Universe, after initial cooling (following the Big Bang), was reheated and reionized by hot stars in newborn galaxies at a redshift in the range $6 < z < 14$ (ref. 1). Though several candidate galaxies at redshift $z > 7$ have been identified photometrically^{2,3}, galaxies with spectroscopically confirmed redshifts have been confined to $z < 6.6$ (refs 4–8). Here we report a spectroscopic redshift of $z = 6.96$ (corresponding to just 750 Myr after the Big Bang) for a galaxy whose spectrum clearly shows Lyman- α emission at 9,682 Å, indicating active star formation at a rate of $\sim 10 M_{\odot} \text{yr}^{-1}$, where $M_{\odot}$ is the mass of the Sun. This demonstrates that galaxy formation was under way when the Universe was only ~ 6 per cent of its present age. The number density of galaxies at $z \approx 7$ seems to be only 18–36 per cent of the density at $z = 6.6$.

Narrowband filter surveys searching for redshifted Lyman- α emitting galaxies (LAEs) through dark atmospheric windows in which the foreground telluric OH band emission is not prohibitively bright have been successful in isolating galaxies at high redshift^{4,5}. Some of the discovered LAEs are gravitationally lensed objects behind clusters of galaxies. To obtain unbiased information about the population of LAEs, systematic surveys at various redshifts have been performed, for example, in the Subaru Deep Field (SDF)^{6–8}. Such studies have been limited to a redshift range of $z < 6.6$. New attempts are in progress^{9–12} to search for galaxy populations at $7 < z < 10$, using deep narrowband imaging observations with near-infrared cameras, aiming for high sensitivity despite the limited survey volume. We have taken another approach to survey $z = 7$ LAEs, specifically using the last OH window accessible with a wide-field charge-coupled device (CCD) camera.

We developed a narrowband filter NB973 for the Subaru Suprime-Cam^{13,14}, with a transmission bandwidth of 200 Å centred at 9,755 Å. To investigate the population of $z = 7.0$ LAEs, we conducted an NB973 imaging survey with 15 h exposure (NB973 < 24.9 mag in 5σ) on the photometric nights of 16 and 17 March 2005. An effective area of 876 arcmin² and co-moving depth of 58 Mpc, corresponding to $z = 6.94$ – 7.11 , led to a total probed volume of $3.2 \times 10^5 \text{ Mpc}^3$. (Throughout, we adopt a cosmology with $\Omega_M = 0.3$, $\Omega_A = 0.7$, and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and refer to magnitudes in the AB system measured with a 2'' aperture unless otherwise specified.) We measured the NB973 fluxes of all detected objects and also measured their broadband fluxes using images previously obtained by the SDF project⁸. The colours of objects were derived from these measurements and compared to the expected colours of a $z = 7.0$ LAE, computed using a stellar population synthesis model¹⁵ and an intergalactic absorption model¹⁶. Of the 41,533 objects detected in NB973, we found only two objects with significant NB973 flux that were not detected in any of the broadband images blueward of

NB973. We considered these to be $z = 7$ LAE candidates (hereafter referred to as IOK-1, the brighter of the two in NB973, and IOK-2). Their images and photometric properties, as well as the candidate selection criteria, are shown in Fig. 1 and Table 1.

To confirm whether these candidates are real LAEs, we carried out follow-up spectroscopy on 14 and 15 May, 1 June 2005, and 24 April 2006, using the Faint Object Camera and Spectrograph (FOCAS)¹⁷. Total exposures were 8.5 h and 3.0 h for IOK-1 and IOK-2, respectively. Figure 2 shows the spectrum of IOK-1 with OH skylines subtracted. We identified this object as a LAE at $z = 6.96$, which makes it, to our knowledge, the most distant galaxy ever spectroscopically confirmed. Our discovery of this galaxy provides direct evidence that at least one galaxy already existed only ~ 750 million years after the Big Bang.

Three-hour integration on IOK-2 did not reveal significant spectral features, and our photometric detection could have been due to noise. There are 10^6 independent 2'' apertures in the 876 arcmin² field, and the probability of noise exceeding 5σ for a gaussian distribution is 3×10^{-7} , so that the expected number of 5σ noise

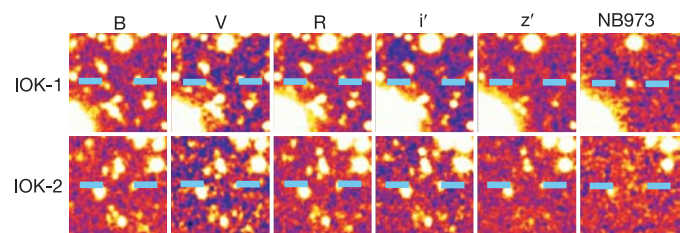


Figure 1 | Multi-waveband $20'' \times 20''$ images of the $z = 6.96$ Lyman α emitter IOK-1 and the unidentified candidate IOK-2. Deep broadband images (Kron–Cousin B, V, R bands, and SDSS i' and z' bands, with limiting magnitudes 28.45, 27.74, 27.80, 27.43 and 26.62 in 3σ , respectively) were taken during 2002–2004 (ref. 8). Dithered NB973 exposures (15–30 min each; 15 h total) were combined to remove fringing produced by foreground sky OH emission lines and yielded a 2'' aperture limiting magnitude of NB973 = 24.9 (5σ). All the images were convolved to a common seeing size of $0.98''$. Of the 41,533 objects detected down to NB973 = 24.9 (total magnitude in this case), we selected those satisfying the following criteria: (1) B, V, R, i' , z' $< 3\sigma$, for $z = 7.0$ LAE candidates IOK-1 and IOK-2, and (2) B, V, R $< 3\sigma$, $i' - z' > 1.3$, $z' - \text{NB973} > 1.0$, for $z = 6.4$ – 7.0 LAE candidate IOK-3. These criteria were determined by model calculations^{15,16}. The $i' - z'$ colour eliminates interlopers such as lower- z [O II], [O III] line-emitting galaxies while the colour $z' - \text{NB973}$ picks up objects with NB973 flux excess with respect to the z' continuum flux. In addition, null ($< 3\sigma$) detection in B, V and R (or B, V, R, i' and z') total magnitudes are required since continuum fluxes blueward of the Lyman- α line should be mostly (or completely) absorbed by the neutral IGM. The criterion-(2) object, IOK-3, was identified as a $z = 6.6$ LAE by an independent spectroscopic survey²² and the possibility of a $z = 7.0$ LAE was excluded.

¹National Astronomical Observatory, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan. ²Department of Astronomy, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

³The Graduate University for Advanced Studies, Mitaka, Tokyo 181-8588, Japan. ⁴Subaru Telescope, 650 North A'ohoku Place, Hilo, Hawaii 96720, USA. ⁵Department of Astronomy, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan. ⁶The Institute of Astronomy, University of Tokyo, Mitaka, Tokyo 181-0015, Japan. ⁷Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA.

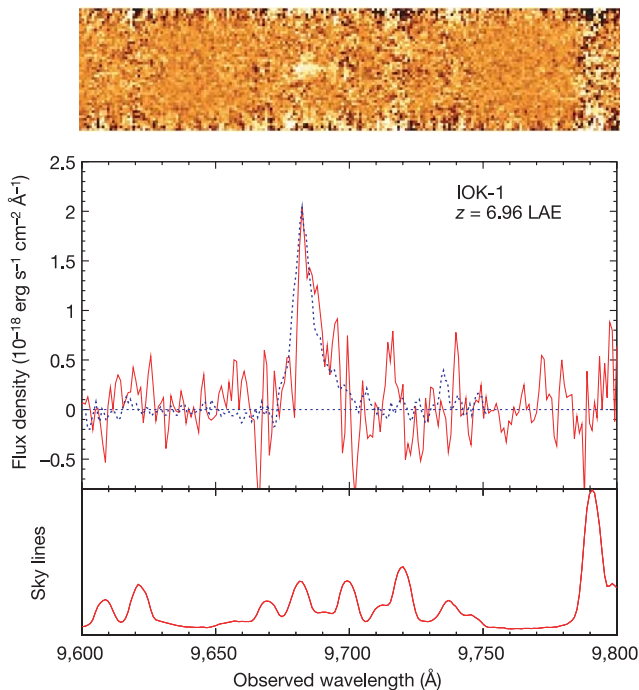


Figure 2 | Combined spectrum of $z = 6.96$ galaxy, IOK-1. The bottom panel shows the removed OH sky emission lines. An echelle grism (175 lines per millimetre, resolution $\approx 1,600$) with z' filter and $0.8''$ slit was used to obtain 11 spectra of 30-min exposure each, dithered along the slit by $\pm 1''$. The Lyman- α emission peak is located at $9,682 \text{ \AA}$. This feature was confirmed to follow the dithered shifts along the slit. The Lyman- α emission has a total flux of $2.0 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$, consistent with the estimate from the NB973 image ($2.7 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$), and a full width at half-maximum of $13 \text{ \AA}$. The Lyman- α luminosity is $L_{\text{Ly}\alpha} = 1.1 \times 10^{43} \text{ erg s}^{-1}$, corresponding to $\text{SFR}_{\text{Ly}\alpha} = 10 M_{\odot} \text{ yr}^{-1}$. The emission feature is significant at the 5.5σ level. The asymmetric emission line profile matches the composite template line profile (dashed line) produced from 12 Lyman- α emitters at $z = 6.6$ (ref. 22) normalized and shifted to $z = 6.96$. This emission cannot be unresolved $z = 1.6$ [O II] doublet lines with $7.0 \text{ \AA}$ separation ($3,726 \text{ \AA}$, $3,729 \text{ \AA}$ at rest frame), as finer sky lines (about $5.1\text{--}6.2 \text{ \AA}$) are clearly resolved. Similarly, the line is not H β , [O III] $4,959 \text{ \AA}$, [O III] $5,007 \text{ \AA}$, H α , [S II] $6,717 \text{ \AA}$ or [S II] $6,731 \text{ \AA}$, because our spectrum shows none of the other lines that should accompany any of these lines in the observed wavelength range. The possibility that OH lines mask other lines was carefully examined for each case and ruled out. There appears to be an extremely weak emission at around $z = 7.02$ in the 3 h integration spectrum of IOK-2, but this needs further integration for confirmation.

events is 0.3. Another possible origin for IOK-2 is a transient object such as a supernova or active galactic nucleus that happened to brighten in March 2005, but was inconspicuous in broadband images taken 1–2 years earlier. Counts of variable objects in i' band images from the SDF taken over 1–2 years indicate that the expected number of variable objects having brightness amplifications large enough to exceed our detection limit (NB973 = 24.9) is at most about three.

Table 1 | Properties of $z \approx 7$ Lyman- α emitter candidates

ID	Position (J2000)	i' (mag)	z' (mag)	NB973 ($2''$)	NB973 (total)	$L_{\text{Ly}\alpha}$ ($10^{43} \text{ erg s}^{-1}$)	$\text{SFR}_{\text{Ly}\alpha}$ ($M_{\odot} \text{ yr}^{-1}$)
IOK-1	$\alpha = 13 \text{ h } 23 \text{ min } 59.8 \text{ s}$, $\delta = +27^{\circ} 24' 55.8''$	>27.84	>27.04	24.60	24.40	1.1 ± 0.2	10 ± 2
IOK-2	$\alpha = 13 \text{ h } 25 \text{ min } 32.9 \text{ s}$, $\delta = +27^{\circ} 31' 44.7''$	>27.84	>27.04	25.51	24.74	$<1.1 \pm 0.2$	$<10 \pm 2$

Magnitudes (SDSS i' , z' filters and NB973) are in the AB system, measured with a $2''$ aperture. The NB973 total magnitudes integrated over the entire image are also given. The Lyman- α luminosity $L_{\text{Ly}\alpha}$ of IOK-1 was derived from spectroscopy, while the upper limit of $L_{\text{Ly}\alpha}$ for IOK-2 was evaluated using NB973 total magnitude by assuming that 68% of the NB973 flux is in the Lyman- α line, on the basis of simulations. Luminosities were then converted into corresponding Lyman- α star-formation rate $\text{SFR}_{\text{Ly}\alpha}$ using the relation derived from Kennicutt's equation²⁷ with case B recombination theory. 1σ errors are also given for $L_{\text{Ly}\alpha}$ and $\text{SFR}_{\text{Ly}\alpha}$. IOK-1 was spectroscopically identified as a $z = 6.96$ Lyman- α emitter. For IOK-2, the photometric estimation of Lyman- α luminosity should be regarded as an upper limit because we were unable to make a spectroscopic confirmation.

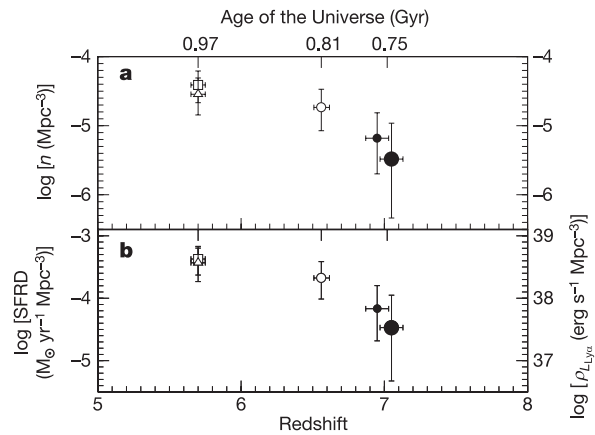


Figure 3 | Decline of the number density of LAEs between $6.6 < z < 7.0$. The number density n (a), luminosity density $\rho_{\text{Ly}\alpha}$ and star-formation-rate density SFRD (b) derived from Lyman- α line fluxes detected in several Suprime-Cam LAE surveys are plotted for the three epochs of $z = 5.7$ (square⁶ and triangle¹⁸; other survey fields), 6.6 (open circle⁷; also SDF), and 7.0 (filled circles: our result) down to our detection limit $L_{\text{Ly}\alpha} = 9.9 \times 10^{42} \text{ erg s}^{-1}$ calculated from NB973 = 24.9 (5σ). The volumes probed by these surveys are of a similar order of magnitude ($\sim 2\text{--}3 \times 10^5 \text{ Mpc}^3$). The large filled circle at $z = 7.0$ represents the values derived for the single LAE discovered in the present study, and the small filled circle those for two LAEs, including the unconfirmed candidate IOK-2. All the vertical error bars in both panels include the errors from both poissonian statistics for small numbers and cosmic variance. The horizontal error bars indicate the surveyed redshift ranges. Note that large and small filled circles are shifted slightly horizontally from $z = 7.0$ to make them easier to see.

Those transient objects, if any existed, would not show conspicuous emission line features in the NB973 waveband.

We note, however, that even for the brighter IOK-1, it was necessary to combine the eleven 30-min exposures with the highest signal-to-noise ratio under the best sky conditions selected from the 17 spectra taken to detect the $z = 6.96$ Lyman- α emission shown in Fig. 2 clearly. Because our spectral detection limit for 3 h did not reach the depth required to deny the presence of weak Lyman- α emission, we retain IOK-2 as another possible but unconfirmed LAE. Therefore, we conclude that the detected number of LAEs at $z = 7.0$ in our survey volume is at least one, and possibly could be two at most.

The observed LAE number density is expected to decline beyond the redshift at which the reionization had completed, because the increasing fraction of intergalactic neutral hydrogen absorbs the Lyman- α photons from young galaxies. However, no significant decline of LAE number densities has been yet observed for $3 < z < 6.6$ (ref. 26). Figure 3 shows the volume densities of the LAE numbers, Lyman- α line luminosity, and star formation rate at three redshifts^{6,7,18} down to our detection limit. These physical parameters are useful for tracing the properties of LAEs in an environment with an increasing fraction of neutral hydrogen in the intergalactic medium (IGM). Errors shown in the figure include (1)

the poissonian error associated with small-number statistics and (2) the cosmic variance uncertainty in galaxy number density due to the nonuniform distribution of galaxies that reflect the large-scale structure of the Universe. We estimated the amount of expected cosmic variance using an analytic cold dark matter model¹⁹ assuming a bias parameter of LAEs, $b = 3.4 \pm 1.8$ (ref. 20), and taking SDF survey volumes and LAE number densities at $z = 6.6$ (ref. 7) and 7.0. The resultant cosmic variance of $z = 6.6$ –7.0 LAEs in the SDF down to our detection limit is at most 25%, which is less than the poissonian error described below.

The $z = 6.6$ LAE survey⁷ of the same field (SDF), but with a smaller volume ($2.2 \times 10^5 \text{ Mpc}^3$), photometrically yielded four bright LAEs down to a comparable luminosity limit. Assuming that the evolution of the LAE population is negligible, we should have $6_{-2.8}^{+3.9}$ LAEs within our survey volume at $z = 7.0$. However, our finding of only one (or at most two) LAE(s) at $z = 7.0$ suggests that the LAE number density is only 18% (or 36%) of that at $z = 6.6$. The confidence level of the difference in the number density between $z = 7.0$ and $z = 6.6$ is 93% (or 85%), taking the poissonian error for small-number statistics²¹. The observed variation may have been due to the evolutionary change in the LAE properties. However, there is no quantitative way to assess this effect reliably for the moment.

The decrease in the LAE number density at $z = 7.0$ compared with $z = 6.6$, if confirmed in other fields and for the fainter end of the population, could reflect a deficit in Lyman- α photons owing to attenuation by the neutral IGM. A similar decrease in the number density of LAEs was also recently found over the redshift range $5.7 < z < 6.6$ (ref. 22). The neutral hydrogen fraction of the IGM, $x_{\text{H I}}$, at $z = 7.0$ is estimated to be about 0.63–0.83, applying the ratio of the luminosity density at $z = 7.0$ to that at $z = 6.6$ to the Lyman- α line attenuation model²³ and assuming complete reionization at $z = 6.6$. Note that our result does not contradict the Gunn–Peterson test²⁴ (diagnostics of intergalactic Lyman- α absorptions on high redshift quasars) that infers the completion of reionization at $z \approx 6.0$ (ref. 25). This is because the LAE number or luminosity density is sensitive to $x_{\text{H I}} \geq 0.1$ (ref. 26), whereas the quasar Gunn–Peterson trough is sensitive to $x_{\text{H I}} \approx 0.01$ (ref. 25). Therefore, it is natural that the LAE density should not change much at $z < 6.6$, when the quasar Gunn–Peterson troughs show significant attenuation of Lyman- α photons.

Note, however, that the present survey is not sufficiently deep to sample the entire LAE population (5σ detection limit $\geq 9.9 \times 10^{42} \text{ erg s}^{-1}$). Follow-up imaging and spectroscopic surveys in the SDF are planned, to probe fainter $z = 7.0$ LAEs by using red-sensitive CCDs to be updated on Suprime-Cam. Independent surveys of other fields are essential for investigating cosmic variance more accurately and for studying field-to-field variations in reionization status.

Received 3 April; accepted 18 July 2006.

1. Page, L. *et al.* Three year Wilkinson Microwave Anisotropy Probe (WMAP) observations: polarization analysis. *Astrophys. J.* (submitted); preprint at (<http://www.arXiv.org/astro-ph/0603450>) (2006).
2. Egami, E. *et al.* Spitzer and Hubble Space Telescope constraints on the physical properties of the $z \sim 7$ galaxy strongly lensed by A2218. *Astrophys. J.* **618**, L5–L8 (2005).
3. Bouwens, R. J. *et al.* Constraints on $z \approx 10$ galaxies from the deepest Hubble Space Telescope NICMOS fields. *Astrophys. J.* **624**, L5–L8 (2005).
4. Hu, E. M., McMahon, R. G. & Cowie, L. L. An extremely luminous galaxy at $z = 5.74$. *Astrophys. J.* **521**, L9–L12 (1999).

5. Hu, E. M. *et al.* A redshift $z = 6.56$ galaxy behind the cluster Abell 370. *Astrophys. J.* **568**, L75–L79 (2002).
6. Ajiki, M. *et al.* A Subaru search for Ly α emitters at redshift 5.7. *Astron. J.* **126**, 2091–2107 (2003).
7. Taniguchi, Y. *et al.* The SUBARU Deep Field Project: Lyman α emitters at a redshift of 6.6. *Publ. Astron. Soc. Jpn* **57**, 165–182 (2005).
8. Kashikawa, N. *et al.* The Subaru Deep Field: The optical imaging data. *Publ. Astron. Soc. Jpn* **56**, 1011–1023 (2004).
9. Stark, D. P. & Ellis, R. S. Searching for the sources responsible for cosmic reionisation: probing the redshift range $7 < z < 10$ and beyond. *New Astron. Rev.* **50**, 46–52 (2006).
10. Willis, J. P. & Courbin, F. A deep, narrow J-band search for protogalactic Ly- α emission at redshift $z \sim 9$. *Mon. Not. R. Astron. Soc.* **357**, 1348–1356 (2005).
11. Barton, E. J., Bullock, J. S., Cooray, A. & Kaplinghat, M. First light and reionisation: a conference summary. *New Astron. Rev.* **50**, 1–12 (2006).
12. Horton, A. *et al.* DAZLE: The Dark Ages z (redshift) Lyman- α explorer. *Proc. SPIE* **5492**, 1022–1032 (2004).
13. Iye, M. *et al.* Current performance and on-going improvements of the 8.2 m Subaru Telescope. *Publ. Astron. Soc. Jpn* **56**, 381–397 (2004).
14. Miyazaki, S. *et al.* Subaru Prime Focus Camera—Suprime-Cam. *Publ. Astron. Soc. Jpn* **54**, 833–853 (2002).
15. Bruzual, G. & Charlot, S. Stellar population synthesis at the resolution of 2003. *Mon. Not. R. Astron. Soc.* **344**, 1000–1028 (2003).
16. Madau, P. Radiative transfer in a clumpy Universe: the colors of high-redshift galaxies. *Astrophys. J.* **441**, 18–27 (1995).
17. Kashikawa, N. *et al.* FOCAS: The Faint Object Camera and Spectrograph for the Subaru Telescope. *Publ. Astron. Soc. Jpn* **54**, 819–832 (2002).
18. Hu, E. M. *et al.* The luminosity function of Ly- α emitters at redshift $z \sim 5.7$. *Astron. J.* **127**, 563–575 (2004).
19. Sheth, R. K. & Tormen, C. Large-scale bias and the peak background split. *Mon. Not. R. Astron. Soc.* **308**, 119–126 (1999).
20. Ouchi, M. *et al.* The discovery of primeval large-scale structures with forming clusters at redshift 6. *Astrophys. J.* **620**, L1–L4 (2005).
21. Gehrels, N. Confidence limits for small numbers of events in astrophysical data. *Astrophys. J.* **303**, 336–346 (1986).
22. Kashikawa, N. *et al.* The end of the reionisation epoch probed by Lyman α emitters at $z = 6.5$ in the Subaru Deep Field. *Astrophys. J.* **648** (in the press); preprint at (<http://www.arXiv.org/astro-ph/0604149>) (2006).
23. Santos, M. R. Probing reionization with Lyman α emission lines. *Mon. Not. R. Astron. Soc.* **349**, 1137–1152 (2004).
24. Gunn, J. E. & Peterson, B. A. On the density of neutral hydrogen in intergalactic space. *Astrophys. J.* **142**, 1633–1641 (1965).
25. Fan, X. *et al.* Evolution of the ionizing background and the epoch of reionisation from the spectra of $z \sim 6$ quasars. *Astron. J.* **123**, 1247–1257 (2002).
26. Malhotra, S. & Rhoads, J. E. Luminosity functions of Ly α emitters at redshifts $z = 6.5$ and $z = 5.7$: evidence against reionisation at $z \leq 6.5$. *Astrophys. J.* **617**, L5–L8 (2004).
27. Kennicutt, R. C. J. Star formation in galaxies along the Hubble sequence. *Annu. Rev. Astron. Astrophys.* **36**, 189–232 (1998).

Acknowledgements This work is based on data collected with the Subaru Telescope, operated by the National Astronomical Observatory of Japan and was partly supported by KAKENHI. We thank the SDF team and the staff at the Subaru Observatory for providing the data and assisting us with observations and data reduction. K.O. acknowledges a fellowship from the Japan Society for the Promotion of Science.

Author Contributions M.I., the principal investigator, proposed this project, designed the NB973 filter and led the observations. K.O. made all the observations, including the slit mask design for the spectroscopy, carried out all the data reduction and analysis and produced the figures. M.I. and K.O. jointly wrote the paper. N.K. contributed to observations, data analysis and discussion. H.F. and Takashi, H. helped with the Suprime-Cam imaging and FOCAS spectroscopy, respectively, as instrument support astronomers. K.S. and M.O. commented on the manuscript. T.M. investigated the possibility of the unconfirmed candidate being a variable object. Y.M., K.O. and Tetsuya H. measured the transmission curve of the NB973 filter.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.I. (iye@optik.mtk.nao.ac.jp).

Rapid evolution of the most luminous galaxies during the first 900 million years

Rychard J. Bouwens¹ & Garth D. Illingworth¹

The first 900 million years (Myr) to redshift $z \approx 6$ (the first seven per cent of the age of the Universe) remains largely unexplored for the formation of galaxies. Large samples of galaxies have been found at $z \approx 6$ (refs 1–4) but detections at earlier times are uncertain and unreliable. It is not at all clear how galaxies built up from the first stars when the Universe was about 300 Myr old ($z \approx 12$ –15) to $z \approx 6$, just 600 Myr later. Here we report the results of a search for galaxies at $z \approx 7$ –8, about 700 Myr after the Big Bang, using the deepest near-infrared and optical images ever taken. Under conservative selection criteria we find only one candidate galaxy at $z \approx 7$ –8, where ten would be expected if there were no evolution in the galaxy population between $z \approx 7$ –8 and $z \approx 6$. Using less conservative criteria, there are four candidates, where 17 would be expected with no evolution. This demonstrates that very luminous galaxies are quite rare 700 Myr after the Big Bang. The simplest explanation is that the Universe is just too young to have built up many luminous galaxies at $z \approx 7$ –8 by the hierarchical merging of small galaxies.

High-redshift galaxies can be identified using the well-established “dropout” technique⁵. Absorption shortward of 91.2 nm and 121.6 nm by intervening neutral hydrogen results in a sharp change (a break) in the spectral energy distribution in redshifted galaxies and can be used to determine their redshift⁶. In the three years since the first galaxies were discovered in the reionization epoch^{7,8} using this dropout technique, 500 galaxies¹ are now known at $z \approx 6$ and the frontier for exploration has shifted to galaxies at $z \approx 7$ –8 and higher redshifts. Finding galaxies at redshifts beyond $z \approx 6$ is very challenging, however, because the sources are extremely faint, and the high sky backgrounds make ground-based searches in the near-infrared part of the spectrum very difficult. Although near-infrared imaging from space largely overcomes this problem, we are limited by the areal coverage of the available instrumentation. For example, the Hubble Space Telescope (HST) Near Infrared Camera and Multi-object Spectrograph (NICMOS) instrument⁹—the only near-infrared imaging instrument in space—has a very small field of view (0.8 arcmin²), which is just 7% of the area of the optical camera on HST, the Advanced Camera for Surveys (ACS). This has made it extremely time-consuming to survey large areas of sky for high-redshift sources.

Nonetheless, we were able to use the deep near-infrared NICMOS data over the Hubble Ultra Deep Field (HUDF)¹⁰ to conduct the first significant search for $z \approx 7$ –8 galaxies^{3,11}. The few candidates found were consistent with a slight increase in the volume density of luminous sources from $z \approx 7$ –8 to $z \approx 6$ (from 700 Myr to 900 Myr after the Big Bang). However, owing to possible instrumental problems for some of the candidates and the large uncertainty caused by large-scale structure effects, we were not able to draw very strong conclusions about the evolution of the luminosity function, the ultraviolet (UV) luminosity density, and the star-formation rate

from $z \approx 7$ –8 to $z \approx 6$. Attempts to establish these quantities based on other data sets have also been quite uncertain¹².

Fortunately, sufficient NICMOS and ACS data are now available that we can firmly establish the density of luminous galaxies at $z \approx 7$ –8 and the evolution with respect to $z \approx 6$. These data come from many years of primary and parallel observations with HST and include $> \sim 20$ arcmin² of deep NICMOS data (reaching J and H band limits of ~ 27 –28 AB mag at 5σ) that overlap with HST ACS data of comparable depth. A substantial fraction of these data are available over the two Great Observatories Origins Deep Survey fields¹³ and permit a much more robust search for $z \approx 7$ –8 galaxies than was possible with the HUDF data alone—significantly increasing our search area and allowing us to average over the effects of large-scale structure. The search fields used for our $z \approx 7$ –8 selection are given in Supplementary Fig. 1 and Supplementary Table 1^{10,13–15}. These extensive ACS and NICMOS data sets were processed using recently developed reduction software¹⁶ (R.J.B. *et al.*, manuscript in preparation). After registration of these data onto the same frame, object catalogues were generated with SExtractor¹⁷ from all 4σ detections found in a χ^2 image (generated from the J and H images). These catalogues were cleaned of spurious or other marginal detections by requiring objects to be 4.5σ in the H band (0.3" diameter aperture) and the photometry performed, after smoothing the images to a consistent point spread function.

We then considered two different z-dropout selections on these catalogues. z-dropout selections identify star-forming galaxies at $z \approx 7$ –8 that show a strong flux break between the z and J bands due to the 121.6 nm absorption feature, but have very blue infrared colours (see Supplementary Fig. 2). Our first selection was a relatively conservative one that was designed to isolate a robust set of z-dropouts. Our second selection was a little more aggressive and designed to find a larger sample of candidates (albeit less secure) and to test the robustness of the results derived from our first selection. Applying these criteria to our photometric catalogues, only one z-dropout candidate made our conservative selection and four candidate z-dropouts made our less-conservative selection. There were two other sources in our search fields that satisfied our selection criteria, but they are very likely to be T-dwarf stars. All four $z \approx 7$ –8 candidates are shown in Fig. 1 and their properties are listed in Supplementary Table 2. A detailed discussion of our current selection is provided in the Supplementary Information.

What do these z-dropout candidates imply for the evolution of the rest-frame UV luminosity function? (The UV luminosity function describes the number density of the galaxies versus the UV luminosity and is usually parameterized as $\phi^* e^{-L/L^*} (L/L^*)^\alpha$ where ϕ^* is the normalization of the galaxy population, L^* is the characteristic luminosity of the galaxy population, and α is the faint-end slope of the galaxy population.) We compared the current numbers with what we might have expected if galaxies at $z \approx 7$ –8 were similar to those at

¹UCO/Lick Observatory and Department of Astronomy, University of California Santa Cruz, Santa Cruz, California 95064, USA.

$z \approx 6$, just 200 million years later. We can compute the expected number by using our cloning software to artificially redshift the $z \approx 6$ population over the redshift range of our sample, add them to our data, and then repeat the selection^{18,19}. We estimated that we would find 10.1 ± 3.5 z-dropouts using our conservative selection and 17.0 ± 4.9 z-dropouts using our less conservative selection. This is clearly substantially more than the 1 and 4 z-dropout candidates found in our conservative and less-conservative selections, respectively, which suggests that there has been substantial evolution for galaxies that lie at the bright end of the luminosity function. For simple poissonian statistics, the present samples are inconsistent with no evolution in the galaxy population between $z \approx 7-8$ and $z \approx 6$, at 99.90% and 99.93% confidence, respectively.

To make a proper assessment of the significance of this result, however, we need to account for the effects of galaxy clustering ('large-scale structure'). Modelling the selection volume in our individual fields with a window of depth $\Delta z = 1$ at $z \approx 7-8$ and bias of 7 (see Supplementary Information for details), we estimate the

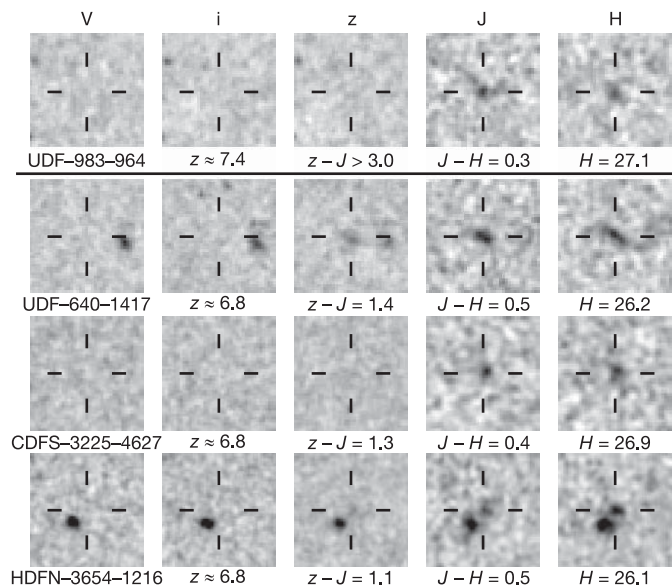


Figure 1 | Optical and near-infrared images of four candidate galaxies at $z \approx 7-8$. The uppermost candidate is the only object satisfying our most conservative z-dropout criterion ($z_{850} - J_{110} > 1.3$, $z_{850} - J_{110} > 1.3 + 0.4(J_{110} - H_{160})$, $J_{110} - H_{160} < 1.2$, undetected [$<2\sigma$] in the V_{606} and i_{775} bands). This object is undetected ($<2\sigma$) in the extremely deep optical B-V-i-z data available in the HUDF, has a $z_{850} - J_{110}$ colour redder than 3.0 mag (1σ), and a $J_{110} - H_{160}$ colour of 0.3 mag. It also shows a significant 5σ detection in the InfraRed Array Channel (IRAC) 3.6- μ m and 4.5- μ m coverage (M. Dickinson *et al.*, manuscript in preparation) that we have from the Spitzer Space Telescope³¹, with optical-infrared colours very similar to dropouts found at $z \approx 6$ (refs 23–26). While this object lacks spectroscopic confirmation (and will probably continue to do so, because it is so faint), it seems extremely probable that it is at $z \approx 7-8$. It had previously been reported as a z-dropout candidate in the HUDF¹¹. The lower three candidates satisfy our less-conservative z-dropout criterion ($z_{850} - J_{110} > 0.8$, $z_{850} - J_{110} > 0.8 + 0.4(J_{110} - H_{160})$, $J_{110} - H_{160} < 1.2$, undetected [$<2\sigma$] in the V_{606} and i_{775} bands). Though each of these sources is consistent with being at the low-redshift edge of the selection ($z \approx 6.8$), the optical imaging available for two of these sources (CDFS-3225-4627 and HDFN-3654-1216) is not deep enough to completely rule out their being low-redshift sources (or T-dwarf stars in the case of CDFS-3225-4627). All z-dropout candidates considered here were detected at $>4\sigma$ and $>4.5\sigma$ in the J and H bands, respectively, and were found in NICMOS data with at least seven individual exposures. The V, i, z, J and H images shown here correspond to the ACS F606W, ACS F775W, ACS F850LP, NICMOS F110W, and NICMOS F160W filters, respectively, with central wavelengths of 591, 776, 944, 1,119 and 1,604 nm, respectively. Each cutout is $3.5'' \times 3.5''$ in size.

field-to-field variance to be $\sim 41\%$ and $\sim 49\%$ for our ~ 5 -arcmin² fields and 0.8-arcmin² fields, respectively^{20,21}. Including the effect of large-scale structure increases the uncertainties on our expected numbers, that is, $10.1^{+3.1}_{-5.0}$ sources with our conservative search and $17.0^{+4.7}_{-7.2}$ sources for our less-conservative search. Again, this is inconsistent with no galaxy evolution, at 99.5% and 99.4% confidence, respectively, and indicates a dramatic drop in the overall prevalence of luminous galaxies over this short time interval. We estimate that the number density of luminous galaxies at $z \approx 7-8$ is just $0.10^{+0.19}_{-0.07}$ times that at $z \approx 6$ for our conservative search and $0.24^{+0.20}_{-0.12}$ times the $z \approx 6$ value for our less-conservative selection. The two determinations here are consistent within the errors and suggest that our basic result is robust. For simplicity, we will use statistics from our more rigorous, conservative search for the remainder of this paper.

What does this deficit of luminous star-forming objects at $z \approx 7-8$ tell us about evolution of the UV luminosity function? This deficit could be due to a change in the overall number density of galaxies or due to a change in luminosity of the brightest objects (or some combination thereof). Perhaps the simplest and most natural explanation, though, is through an evolution of the characteristic luminosity L^* . It is already well established that L^* brightens significantly over the interval $z \approx 6$ to $z \approx 3$ (Fig. 2)^{1,2}, so it seems reasonable to

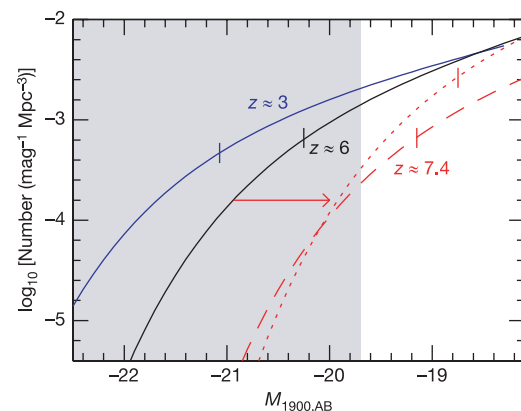


Figure 2 | UV luminosity functions at redshift $z \approx 3, 6$ and 7.4 . The red lines show two possible luminosity functions at $z \approx 7-8$ that are consistent with the surface density of $z \approx 7-8$ galaxies (z-dropouts). These luminosity functions are compared with existing $z \approx 6$ and $z \approx 3$ luminosity functions^{1,29}. The characteristic luminosity for each luminosity function is indicated on the figure with a short vertical line. The grey region shows the approximate flux range of the current z-dropout study. The luminosity functions here are plotted in terms of the number of galaxies per magnitude per cubic megaparsec and are a function of the absolute magnitude of galaxies (M_{AB}) at 1900 Å. The rest-frame UV luminosity functions brightens considerably from $z \approx 6$ to $z \approx 3$ (refs 1, 2), and it seems reasonable to imagine that this trend continues from even earlier times (that is, from $z \approx 7-8$ to $z \approx 6$). The first luminosity function (dashed line) was obtained by requiring the $z \approx 7-8$ luminosity function to have the same normalization ϕ^* and faint-end slope α as the luminosity function at $z \approx 6$ (solid black line)¹, but changing the characteristic luminosity L^* to match the observed surface density of $z \approx 7-8$ galaxies. The second luminosity function (dotted line) was obtained by requiring the product of ϕ^* and L^* to be a constant (to maintain an approximately constant luminosity density), fixing α to the $z \approx 6$ value ($\alpha = -1.73$), but changing the characteristic luminosity L^* to match the observed surface density of $z \approx 7-8$ galaxies. The characteristic luminosity which best fits our search results is 1.1 ± 0.4 mag fainter than the $z \approx 6$ value for the first luminosity function and 1.5 ± 0.4 mag fainter for the second luminosity function (equivalent to $H_{160,AB}$ -band magnitudes of 28.0 ± 0.4 mag and 28.4 ± 0.4 mag, respectively). This best-fit characteristic luminosity is 0.1 mag fainter and 0.1 mag brighter assuming faint-end slopes α of -1.4 and -2.0 , respectively. The 'concordance' cosmology ($\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 \approx 70$ km s⁻¹ Mpc⁻¹) is assumed in this figure and in Fig. 3.

imagine that this trend continues from $z \approx 7-8$ to $z \approx 6$. A relevant question to ask is how much fainter the characteristic luminosity at $z \approx 7-8$ would need to be to reproduce the observed decline. There are two obvious ways to model this change: (1) vary L^* while keeping the normalization $\phi^* = 0.002 \text{ Mpc}^{-3}$ and faint-end slope $\alpha = -1.73$ fixed at the $z \approx 6$ values and (2) vary L^* while keeping the product of ϕ^* and L^* fixed (and α fixed) to maintain a constant luminosity density at $z \geq 6$. The latter model is motivated by a number of recent results at high redshift, which suggest that the rest-frame UV luminosity density—and the total star-formation rate density—shows very mild evolution^{1,22}. These latter observations include the high stellar masses found in many $z \approx 5-6$ systems and large optical depths measured by the Wilkinson Microwave Anisotropy Probe ($\tau = 0.09 \pm 0.03$), both of which suggest substantial high-redshift ($z > 6$) star formation²³⁻²⁸. For the above models, we can evaluate the evolution in the characteristic luminosity needed to

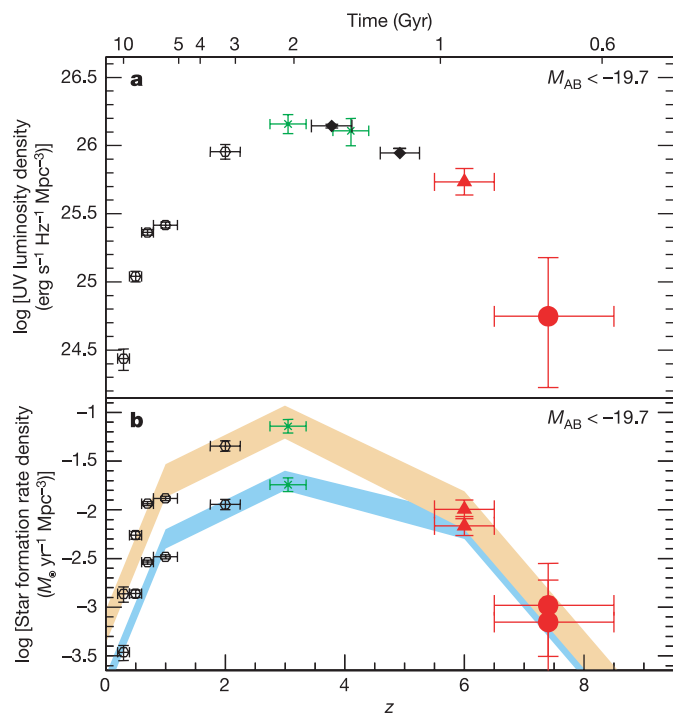


Figure 3 | The luminosity density and star-formation rate density of the Universe over time. **a**, The UV luminosity density at $1,900 \text{ \AA}$ from $z \approx 0$ to $z \approx 7-8$. Because of the flux limits of the current z -dropout searches, we plot the luminosity densities only down to a flux limit of $M_{\text{AB}} \approx -19.7$ mag (which is equivalent to 0.6 times the characteristic luminosity at $z = 6$)¹. The new determination is shown as a large filled circle at $z \approx 7.4$ (with 1σ errors that include both poissonian and large-scale structure uncertainties) and compared with determinations at $z \approx 0-2$ (open hexagons)³⁰, $z \approx 3$ (green crosses)²⁹, $z \approx 4-5$ (black diamonds)²², and $z \approx 6$ (red triangle)¹. Vertical error bars on the latter determinations are all 1σ while the horizontal error bars show the approximate redshift range of applicability. The luminosity density shows an abrupt increase from $z \approx 7-8$ to $z \approx 6$, indicating that there has been a rapid build-up in luminous galaxies (those >0.6 times the characteristic luminosity at $z = 6$, that is, $>0.6L^*_{z=6}$) over this range in redshift—which corresponds to a period of just 200 million years. **b**, The star-formation-rate history of the Universe shown with and without a correction for dust extinction (upper and lower points, respectively). The errors on the points are the same as in **a**. This star-formation-rate history is plotted in terms of the number of solar-mass stars forming per year per unit comoving megaparsec and is derived here from the UV luminosity densities assuming a Salpeter initial mass function, which extends from $0.1 M_{\odot}$ to $100 M_{\odot}$, where $M_{\odot}$ is the mass of the Sun³². To account for the apparent evolution in the UV colours and thus the dust properties of high-redshift galaxies^{1,26,33}, we apply more minimal (~ 0.4 mag) dust corrections at $z > 5$, but revert to more standard values (~ 1.4 mag) at $z \leq 3$ (ref. 30).

produce the factor-of-ten deficit in luminous galaxies at $z \approx 7-8$ by repeating the Monte Carlo simulations described above. The end result is that we find that the characteristic luminosity at $z \approx 7-8$ in the rest-frame UV ($\sim 1,900 \text{ \AA}$) must be 1.1 ± 0.4 mag fainter than $z \approx 6$ to reproduce the observed result for our first set of assumptions and 1.5 ± 0.4 mag fainter for our second (see Fig. 2). In both cases, the evolution is fairly dramatic and suggests that most galaxies at $z \approx 7-8$ are much fainter than our survey limits.

The evolution observed here can best be presented in the terms of the rest-frame continuum UV luminosity density integrated down to our approximate magnitude limit $M_{\text{AB}} \approx -19.7$ mag and compared with similar determinations at other redshifts (Fig. 3)^{1,22,29,30}. As is evident from Fig. 3, the luminosity density, and especially the star-formation rate (the extinction-corrected luminosity density), increases rapidly at early times and indicates that there is a rapid build-up in the most-luminous systems that continues from $z \approx 7-8$ (700 Myr) to $z \approx 4$ (1,600 Myr).

The simplest explanation for this large deficit of the most luminous galaxies at $z \approx 7-8$ relative to that at $z \approx 6$, just 200 Myr later, is that the Universe is simply too young to have built up many luminous systems, and that the majority of galaxies at $z \approx 7-8$ have luminosities much fainter than $M_{\text{AB}} \approx -19.7$ mag. Hierarchical build-up would suggest such a result, but detailed modelling is needed to establish whether the rate of change measured here is consistent. Determining the nature of the $z \approx 7-8$ star-forming population will be a challenge, requiring significantly deeper observations. The HST Wide-Field Camera 3 will provide a near-term opportunity, but ultimate elucidation of the galaxy buildup in the first 500 Myr awaits the James Webb Space Telescope.

Received 6 April; accepted 1 August 2006.

- Bouwens, R. J., Illingworth, G. D., Blakeslee, J. P. & Franx, M. Galaxies at $z \approx 6$: The UV luminosity function and luminosity density from 506 UDF, UDF-Ps, and GOODS i-dropouts. *Astrophys. J.* (in the press); preprint at (<http://www.arXiv.org/astro-ph/0509641>) (2005).
- Dickinson, M. et al. Color-selected galaxies at $z \approx 6$ in the great observatories origins deep survey. *Astrophys. J.* **600**, 99–102 (2004).
- Yan, H. J. & Windhorst, R. A. Candidates of $z \approx 5.5-7$ galaxies in the Hubble Space Telescope Ultra Deep Field. *Astrophys. J.* **612**, 93–96 (2004).
- Bunker, A. J., Stanway, E. R., Ellis, R. S. & McMahon, R. G. The star formation rate of the Universe at $z \approx 6$ from the Hubble Ultra Deep Field. *Mon. Not. R. Astron. Soc.* **355**, 374–384 (2004).
- Steidel, C. C., Giallisco, M., Pettini, M., Dickinson, M. & Adelberger, K. L. Spectroscopic confirmation of a population of normal star-forming galaxies at redshifts $Z > 3$. *Astrophys. J.* **462**, 17–20 (1996).
- Madau, P. Radiative transfer in a clumpy universe: The colors of high-redshift galaxies. *Astrophys. J.* **441**, 18–27 (1995).
- Becker, R. H. et al. Evidence for reionization at $z \approx 6$: detection of a Gunn-Peterson trough in a $z = 6.28$ quasar. *Astron. J.* **122**, 2850–2857 (2001).
- Fan, X. et al. Evolution of the ionizing background and the epoch of reionization from the spectra of $z \approx 6$ quasars. *Astronom. J.* **123**, 1247–1257 (2002).
- Skinner, C. J. et al. On-orbit properties of the NICMOS detectors on HST. *Proc. SPIE* **3354**, 2–13 (1998).
- Thompson, R. I. et al. The Near-Infrared Camera and Multi-Object Spectrometer Ultra Deep Field: observations, data reduction, and galaxy photometry. *Astron. J.* **130**, 1–12 (2005).
- Bouwens, R. J. et al. Galaxies at $z \approx 7-8$: z_{850} -dropouts in the Hubble Ultra Deep Field. *Astrophys. J.* **616**, 79–82 (2004).
- Richard, J., Pello, R., Schaerer, D., Le Borgne, J.-F. & Kneib, J.-P. Constraining the population of $6 < z < 10$ star-forming galaxies with deep near-IR images of lensing clusters. *Astron. Astrophys.* (in the press); preprint at (<http://www.arXiv.org/astro-ph/0606134>) (2006).
- Giallisco, M. et al. The Great Observatories Origins Deep Survey: initial results from optical and near-infrared imaging. *Astrophys. J.* **600**, 99–102 (2004).
- Thompson, R. I. et al. Near-Infrared Camera and Multi-Object Spectroscopic observations of the Hubble Deep Field: observations, data reduction, and galaxy photometry. *Astron. J.* **117**, 17–39 (1999).
- Dickinson, M. A. Complete NICMOS map of the Hubble Deep Field North. *AIP Conf. Proc.* **470**, 122–132 (1999).
- Blakeslee, J. P., Anderson, K. R., Meurer, G. R., Benítez, N. & Magee, D. An automatic image reduction pipeline for the Advanced Camera for Surveys. in *Astronomical Data Analysis Software and Systems XII* (eds Payne, H. E., Jedrzejewski, R. I. & Hook, R. N.) 257–260 (ASP Conf. Ser. 295, Astronomical Society of the Pacific, 2003).

17. Bertin, E. & Arnouts, S. SExtractor: software for source extraction. *Astron. Astrophys.* **117** (Suppl.), 393–404 (1996).
18. Bouwens, R. J., Broadhurst, T. J. & Silk, J. Cloning Hubble Deep Fields. I. A model-independent measurement of galaxy evolution. *Astrophys. J.* **506**, 557–578 (1998).
19. Bouwens, R. J., Broadhurst, T. J. & Illingworth, G. D. Cloning dropouts: implications for galaxy evolution at high redshift. *Astrophys. J.* **593**, 640–660 (2003).
20. Mo, H. J. & White, S. D. M. An analytic model for the spatial clustering of dark matter haloes. *Mon. Not. R. Astron. Soc.* **282**, 347–361 (1996).
21. Somerville, R. S. *et al.* Cosmic variance in the Great Observatories Origins Deep Survey. *Astrophys. J.* **600**, 171–174 (2004).
22. Giavalisco, M. *et al.* The rest-frame ultraviolet luminosity density of star-forming galaxies at redshifts $z > 3.5$. *Astrophys. J.* **600**, 103–106 (2004).
23. Eyles, L. *et al.* Spitzer imaging of i'-drop galaxies: old stars at $z \sim 6$. *Mon. Not. R. Astron. Soc.* **364**, 443–454 (2005).
24. Yan, H. *et al.* Rest-frame ultraviolet-to-optical properties of galaxies at $z \sim 6$ and 5 in the Hubble Ultra Deep Field: from Hubble to Spitzer. *Astrophys. J.* **634**, 109–127 (2005).
25. Dow-Hygelund, C. C. *et al.* UV continuum spectroscopy of a $6L^* z = 5.5$ galaxy. *Astrophys. J.* **630**, 137–140 (2005).
26. Yan, H. J. *et al.* The stellar masses and star formation histories of galaxies at $z \sim 6$: constraints from Spitzer observations in the Great Observatories Origins Deep Survey. *Astrophys. J.* (in the press); preprint at (<http://www.arXiv.org/astro-ph/0604554>) (2006).
27. Stark, D., Bunker, A. J., Ellis, R. S., Eyles, L. P. & Lacy, M. A new measurement of the stellar mass density at $z \sim 5$: implications for the sources of cosmic reionization. *Astrophys. J.* (submitted); preprint at (<http://www.arXiv.org/astro-ph/0604250>) (2006).
28. Spergel, D. *et al.* Wilkinson Microwave Anisotropy Probe (WMAP) three year results: implications for cosmology. *Astrophys. J.* (submitted); preprint at (<http://www.arXiv.org/astro-ph/0603449>) (2006).
29. Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M. & Pettini, M. Lyman-break galaxies at $z > \sim 4$ and the evolution of the ultraviolet luminosity density at high redshift. *Astrophys. J.* **519**, 1–17 (1999).
30. Schiminovich, D. *et al.* The GALEX-VVDS measurement of the evolution of the far-ultraviolet luminosity density and the cosmic star formation rate. *Astrophys. J.* **619**, 47–50 (2005).
31. Labbé, I., Bouwens, R. J., Illingworth, G. D. & Franx, M. Confirmation of z-dropout galaxies in the HUDF with Spitzer IRAC imaging: stellar masses and ages. *Astrophys. J.* (in the press); preprint at (<http://www.arXiv.org/astro-ph/0608444>) (2006).
32. Madau, P., Pozzetti, L. & Dickinson, M. The star formation history of field galaxies. *Astrophys. J.* **498**, 106–116 (1998).
33. Stanway, E. R., McMahon, R. G. & Bunker, A. J. Near-infrared properties of i-drop galaxies in the Hubble Ultra Deep Field. *Mon. Not. R. Astron. Soc.* **359**, 1184–1192 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to L. Bergeron, S. Kassin, D. Magee, M. Stiavelli, R. Thompson and A. Zirm for their help in reducing the NICMOS data, to M. Franx for critical reading of this manuscript, to S. Malhotra for making the NICMOS parallels to her PEARS program public early, and to I. Labbé for assistance in interpreting the IRAC data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.J.B. (bouwens@ucolick.org).

Cooling a nanomechanical resonator with quantum back-action

A. Naik^{1,2}, O. Buu^{1,3}, M. D. LaHaye^{1,3}, A. D. Armour⁴, A. A. Clerk⁵, M. P. Blencowe⁶ & K. C. Schwab¹

Quantum mechanics demands that the act of measurement must affect the measured object. When a linear amplifier is used to continuously monitor the position of an object, the Heisenberg uncertainty relationship requires that the object be driven by force impulses, called back-action^{1–3}. Here we measure the back-action of a superconducting single-electron transistor (SSET) on a radio-frequency nanomechanical resonator. The conductance of the SSET, which is capacitively coupled to the resonator, provides a sensitive probe of the latter's position; back-action effects manifest themselves as an effective thermal bath, the properties of which depend sensitively on SSET bias conditions. Surprisingly, when the SSET is biased near a transport resonance, we observe cooling of the nanomechanical mode from 550 mK to 300 mK—an effect that is analogous to laser cooling in atomic physics. Our measurements have implications for nanomechanical readout of quantum information devices and the limits of ultrasensitive force microscopy (such as single-nuclear-spin magnetic resonance force microscopy). Furthermore, we anticipate the use of these back-action effects to prepare ultracold and quantum states of mechanical structures, which would not be accessible with existing technology.

Back-action impulses arise in practice from the quantized and stochastic nature of the fundamental particles used in the measuring device. For example, in high precision optical interferometers such as the LIGO gravitational wave detector⁴ or in the single-spin force microscope⁵, the position of a test mass is monitored by reflecting laser light off the measured object and interfering this light with a reference beam at a detector. The measured signal is the arrival rate of photons, and one might say that the optical 'conductance' of the interferometer is modulated by the position of the measured object. Back-action forces that stochastically drive the measured object result from the random impact and momentum transfer of the discrete photons. This mechanical effect of light is thought to provide the ultimate limit to the position and force sensitivity of an optical interferometer. Although this photon 'ponderomotive' noise has not yet been detected during the measurement of a macroscopic object⁶, these back-action effects are clearly observed and carefully used in the cooling of dilute atomic vapours to nanokelvin temperatures.

In the experiments reported here, we study an SSET that is capacitively coupled to a voltage-biased (V_{NR}), doubly clamped nanomechanical resonator (Fig. 1). Like the interferometer, the conductance of the SSET is a very sensitive probe of the resonator's position, whereas the particles transported in this case are a mixture of single and Cooper-paired electrons. We have recently shown the SSET to be nearly a quantum-limited position detector⁷; however, reaching the best sensitivity will ultimately be limited by the back-action of the charged particles³, which could not be observed in previous experiments because of insufficient SSET–resonator coupling.

The back-action force of the SSET results in three measurable effects on the resonator: a frequency shift, a damping rate, and position fluctuations. The frequency shift and damping rate are caused by the in-phase and small out-of phase response in the average electrostatic force between the SSET and resonator, as the resonator oscillates. Position fluctuations arise from fluctuations in this force. As electrons and Cooper pairs are transported through the SSET island, the island charge changes stochastically by ± 1 or 2 electrons, causing the electrostatic force to fluctuate with an amplitude of $10^{-13} N_{RMS}$, with a white spectral density of $S_F^{1/2} \approx 10^{-18} N Hz^{-1/2}$, extending to ~ 20 GHz (assuming $V_{NR} = 2$ V and parameters given in Supplementary Information).

These back-action effects are observed by measuring the SSET charge noise in the vicinity of the resonator frequency (Supplementary Information), thus obtaining the resonator noise temperature,

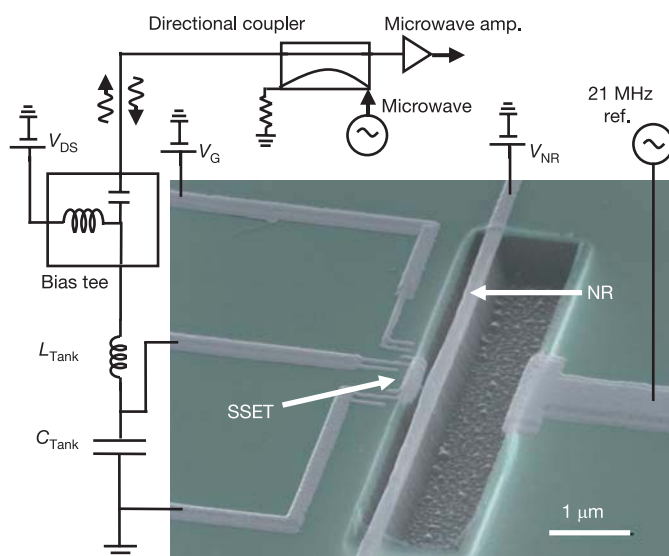


Figure 1 | Nanodevice and measurement diagram. Scanning electron microscope image of the device (false colour): a 21.9-MHz, doubly clamped, SiN and Al nanomechanical resonator (NR) coupled to a superconducting single-electron transistor (SSET), with simplified measurement circuit diagram. See Supplementary Information for sample parameters. V_G , gate voltage; V_{NR} , nanoresonator bias voltage; V_{DS} , drain–source voltage, applied through the bias tee. The SSET is read out by applying 1.17 GHz microwaves through the directional coupler, and monitoring the signal reflected off the LCR tank circuit formed by L_{Tank} , C_{Tank} and the SSET. The charge sensitivity of the system is calibrated by a fixed 21 MHz reference voltage signal applied to a nearby gate capacitor.

¹Laboratory for Physical Sciences, ²Department of Electrical and Computer Engineering, ³Department of Physics, University of Maryland, College Park, Maryland 20740, USA.

⁴School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, UK. ⁵Department of Physics, McGill University, Montreal, QC Canada H3A 2T8.

⁶Department of Physics and Astronomy, Dartmouth College, Hanover, New Hampshire 03755, USA.

T_{NR} , and dissipation rate, $\gamma_{\text{NR}} = \omega_{\text{NR}}/Q_{\text{NR}}$, for different coupling voltages, V_{NR} , and refrigerator temperatures, T_{Bath} ; here ω_{NR} and Q_{NR} are the resonator angular frequency and quality factor. Figure 2 inset shows an example of the measured spectral noise power where the mechanical resonance at $\omega_{\text{NR}} = 2\pi 21.837$ MHz is clearly visible, and accurately fits a simple harmonic oscillator response function, on top of a white power spectrum due to an ultralow-noise microwave preamplifier used to read out the SSET with microwave reflectometry⁸.

For low SSET–nanoresonator coupling strengths, and with the SSET biased close to the Josephson quasiparticle peak⁹ (JQP), T_{NR} simply follows T_{Bath} , with the lowest temperature point corresponding to occupation number $N_{\text{TH}} \approx 25$, and the nanoresonator shows a very low dissipation rate ($Q_{\text{NR}} = 120,000$ at $T_{\text{Bath}} = 100$ mK); see Fig. 2. This is an elementary demonstration of the equipartition theorem and of nanomechanical noise thermometry⁷.

As the SSET bias point is held fixed at a JQP and the coupling voltage is increased, we find clear signatures of all three expected back-action effects. Figure 3 shows the change in resonance frequency and dissipation rate, γ_{NR} , caused by coupling to the SSET. The back-action fluctuation effects on T_{NR} first become noticeable for $T_{\text{Bath}} < 200$ mK (Fig. 2), where we observe an increase and ultimately a saturation of $T_{\text{NR}} \approx 200$ mK; the SSET back-action exerts a stochastic force drive, in excess of the thermal noise of T_{Bath} .

At our highest coupling strength, $V_{\text{NR}} = 15$ V, we find that T_{NR} has a much weaker dependence on T_{Bath} , tending towards a fixed value of ~ 200 mK. In this high coupling limit and for $T_{\text{Bath}} > 200$ mK, we observe the counter-intuitive effect that a noisy, non-equilibrium measuring device can cool the nanomechanical mode.

This behaviour can be understood through models based on a quantum noise approach¹⁰ or a master equation description of the coupled electro-mechanical system¹¹. These models show that the back-action effects are that of an effective thermal bath; similar results have been obtained for oscillators coupled to other non-equilibrium devices: a normal state single-electron transistor^{12,13}, a quantum point contact¹⁴, and in experiments with a ping-pong ball in turbulent air flow¹⁵. The resulting damping and temperature of the

resonator due to coupling to both refrigerator thermal bath (T_{Bath}) and SSET effective bath (T_{SSET}) through the associated damping rates (γ_{Bath} and γ_{SSET}) are then given by:

$$\gamma_{\text{NR}} = \gamma_{\text{Bath}} + \gamma_{\text{SSET}} \quad (1a)$$

$$\gamma_{\text{NR}} T_{\text{NR}} = \gamma_{\text{Bath}} T_{\text{Bath}} + \gamma_{\text{SSET}} T_{\text{SSET}} \quad (1b)$$

As long as γ_{NR} is positive, equations (1a, 1b) provide a simple interpretation of the data in Fig. 2: as the coupling to the SSET is increased, the damping γ_{SSET} increases, so the effective temperature of the resonator is pulled away from T_{Bath} and closer to T_{SSET} . Experimentally, T_{SSET} is easily determined from the value of T_{Bath} where all the curves cross, or the value of T_{NR} when $T_{\text{Bath}} \rightarrow 0$: we find $T_{\text{SSET}} \approx 200$ mK. The curvature observed in T_{NR} at high couplings is probably due to the temperature dependence of γ_{Bath} . Note that T_{SSET} is not determined by the thermodynamic temperature of the carriers in the SSET, but is a measure of the intensity of the charge fluctuations on the SSET.

Using the quantum noise approach^{3,16}, we find that the back-action effects are determined by the asymmetric-in-frequency quantum noise spectrum of the back-action force, $S_F(\omega)$:

$$2m\hbar\omega_{\text{NR}}\gamma_{\text{SSET}} = S_F(\omega_{\text{NR}}) - S_F(-\omega_{\text{NR}}) \quad (2a)$$

$$4mk_{\text{B}}T_{\text{SSET}}\gamma_{\text{SSET}} = S_F(\omega_{\text{NR}}) + S_F(-\omega_{\text{NR}}) \quad (2b)$$

where k_{B} is Boltzmann's constant and m is the oscillator mass, and the positive (negative) frequency noise power describes the noise responsible for the rates of energy emission (absorption) from the nanoresonator to the SSET. Thus our nanomechanical resonator provides a frequency-resolved measurement of the asymmetric,

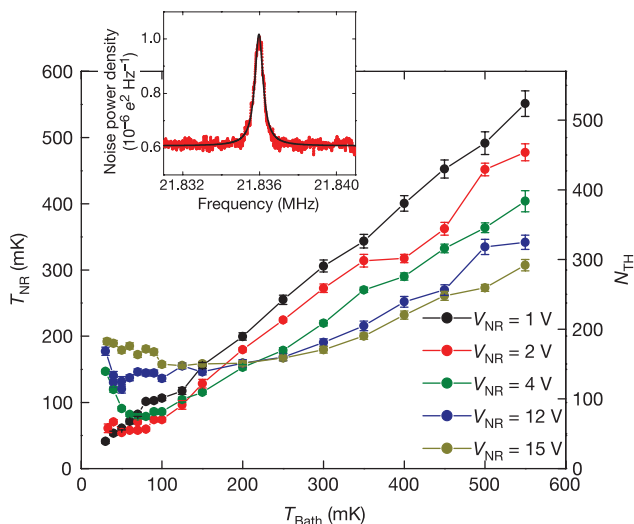


Figure 2 | Resonator temperature versus bath temperature and coupling voltage. Main plot shows the effective temperature of the resonator (in mK and quanta N_{TH}), T_{NR} , versus bath temperature, T_{Bath} , with the SSET biased near the JQP. See text for discussion. Inset shows an example of the measured charge noise power (red trace), taken at $T_{\text{Bath}} = 100$ mK and $V_{\text{NR}} = 4$ V, and a Lorentzian fit (black line). The area of this peak is proportional to the resonator temperature, T_{NR} . Error bars are the standard deviation of the estimate of the total NR noise power.

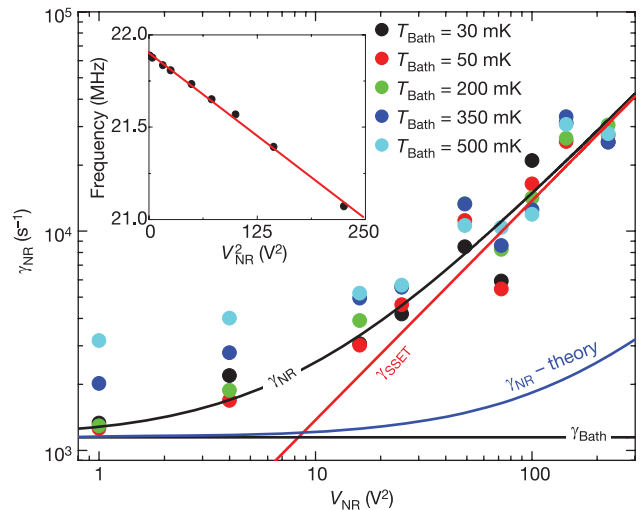
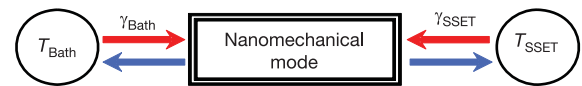


Figure 3 | Resonator damping rate and frequency shift versus coupling voltage. Main plot shows the measured nanoresonator dissipation rate, γ_{NR} , versus V_{NR} for various values of T_{Bath} . Inset shows the resonant frequency versus V_{NR}^2 ($T_{\text{Bath}} = 30$ mK). Both the frequency shift and damping rate to the SSET, γ_{SSET} , scale as V_{NR}^2 , as expected (solid red lines). At low V_{NR} , γ_{NR} is asymptotic to the damping rate to the bath, γ_{Bath} (black line). At high V_{NR} , γ_{NR} is asymptotic to a quadratic dependence, γ_{SSET} although at a rate 14 times higher than expected ($\gamma_{\text{NR}} - \text{theory}$, blue curve.) Black curve shows $\gamma_{\text{NR}} = \gamma_{\text{Bath}} + \gamma_{\text{SSET}}$. The figure above shows the nanomechanical mode coupled to the thermal baths, T_{Bath} and T_{SSET} through dissipative links, γ_{Bath} and γ_{SSET} . The red (blue) arrows indicate the emission (absorption) of quanta by the baths leading to heating (cooling) of the resonator.

quantum noise properties of the SSET^{17–19}, albeit only at the resonator frequency.

When biased at a JQP, the transport is resonant and involves a process of Josephson and quasi-particle tunnelling. This leads to a simple expression for T_{SSET} :

$$k_{\text{B}}T_{\text{SSET}} = \frac{\hbar \Gamma^2 + 4(\Delta E/\hbar)^2}{4 \Delta E/\hbar} \quad (3)$$

where ΔE is the difference between final and initial energies of the tunnelling Cooper pair, and Γ is the quasiparticle tunnel rate, which is essentially temperature independent far below the superconducting transition temperature. This result is analogous to laser ponderomotive cooling of a mechanical cavity²⁰, or to Doppler cooling of a two-level atom²¹; hence we call this ‘Cooper-pair molasses’.

Figure 4 shows the sensitive dependence of T_{NR} , γ_{NR} and ω_{NR} , as we tune the SSET bias point through the two neighbouring JQP resonances. Between the two JQP peaks where $\Delta E > 0$ (‘red detuned’), T_{NR} is in excellent agreement with the theoretical predictions (~ 220 mK) (Fig. 4d). The changes in frequency (Fig. 4b) are also reasonably well described, indicating that the overall magnitude of the non-fluctuating component of the back-action force is close to expectation. Only qualitative agreement is achieved for the damping, which results from a very small asymmetry in the force noise power (equation (2a)). We observe the expected quadratic dependence on

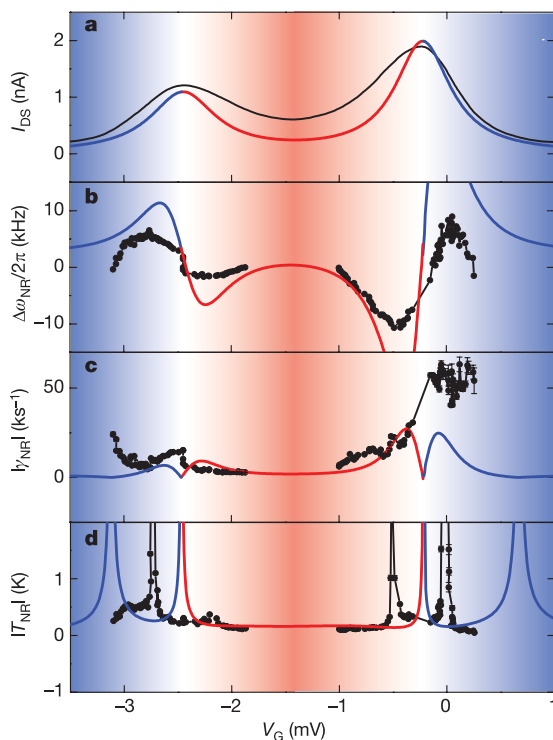


Figure 4 | Back-action effects versus SSET bias point. Comparison of experimental measurements (black points and lines) and theoretical predictions (blue and red lines) as V_{G} is scanned through neighbouring JQP resonances, with fixed $V_{\text{DS}} = 4E_{\text{C}} - 100 \mu\text{V}$, and $V_{\text{NR}} = 5$ V, where E_{C} is the Coulomb energy of the SSET. The panels show **a**, the SSET current, **b**, the resonator frequency shift, **c**, the resonator damping rate, γ_{NR} , and **d**, the effective resonator temperature, T_{NR} . Note that in **c** the theoretically calculated values have been multiplied by a factor of 14. In **c** and **d**, expected unstable bias points, $\{T_{\text{NR}}, \gamma_{\text{NR}}\} < 0$, are shown in blue where SSET is ‘blue detuned’, and stable bias points, $\{T_{\text{NR}}, \gamma_{\text{NR}}\} > 0$, are shown in red where SSET is ‘red detuned’. Error bars are the standard deviation of the fitting parameters to the measured noise spectra.

V_{NR} (Fig. 3) and the increase in damping as one approaches the JQP resonance (Fig. 4c). However, the magnitude is 14 times larger than expected. We believe that this excess may be related to the long-standing discrepancy between theory and experiment in transport characteristics at the JQP²². We do not believe that the excess damping observed is due to another independent bath, as the dissipation is clearly very sensitively controlled by the SSET bias point (Fig. 4).

Outside the JQP resonances, the tunnelling Cooper pairs can emit energy to the resonator ($\Delta E < 0$, ‘blue-detuned’), and the linear theory predicts negative SSET temperatures and damping rates^{10,11}. When the predicted γ_{NR} becomes negative (region shaded blue in Fig. 4), states of small amplitude of motion are unstable with respect to a state of large amplitude motion of the resonator, where the amplitude is ultimately limited by nonlinearity in the dynamics¹⁰, and the picture of the SSET as an effective thermal bath breaks down. A theoretical description of our system in these parameter regions is beyond the scope of this Letter, but we do find experimentally points of dramatically increased T_{NR} outside the JQP peaks, as one would expect qualitatively. Additionally, we have observed the unstable behaviour when biasing the SSET at the double JQP, where the back-action effects are predicted to be larger than at the JQP^{10,11}. Although we are only able to make quantitative comparison for the regions of stable bias, the fact that SSET operating points with opposite sign ΔE yield the same current, but show very different back-action, provides compelling evidence that the SSET is the controlling source of the back-action fluctuations.

We can use equation (2) to calculate the amplitude of the back-action impulses and compare this to the uncertainty principle requirement, $(S_x S_F)^{1/2} \geq \hbar/2$, where S_x is the spectral power density of position (at our highest coupling voltage, we achieve $\sqrt{S_x} = 3 \times 10^{-16} \text{ m Hz}^{-1/2}$). Using $S_x = S_I/(dI/dx)^2$, and $S_I = 2eI$ at the JQP resonance²³, we find $\sqrt{S_F S_x} = 15\hbar/2$, a factor of 15 larger than the uncertainty principle requirement. This excess back-action will limit the position resolution Δx to a factor of 3.9 over the quantum limit (Supplementary Information). However, this level of ideality is already sufficient for an SSET to enable the realization of novel quantum states of the mechanical resonator, including squeezed²⁴ and entangled states²⁵.

The nanomechanical cooling and heating effects reported here are closely related to proposed processes using a Cooper-pair qubit^{26,27} and a quantum dot^{28,29}, and are (to our knowledge) the first examples of the interaction between coherent electronics and nanomechanical modes. Although the cooling power of our Cooper-pair molasses is exceedingly small, $\dot{Q} \approx \gamma_{\text{SSET}} k_{\text{B}}(T_{\text{Bath}} - T_{\text{SSET}}) = 10^{-21}$ W, it is sufficient to cool a single, well isolated, mechanical degree of freedom. Furthermore, we believe that it is possible to use this method to produce ultracold, quantum states in nanomechanical systems with high resistance tunnel junctions, as it is expected¹⁰ that $T_{\text{SSET}} \propto R_{\text{J}}^{-1}$ and $\gamma_{\text{SSET}} \propto R_{\text{J}}^2$, providing increasing coupling to a decreasing temperature bath as the junction resistance, R_{J} , is increased. For instance, the junctions reported in ref. 30 should reach a minimum $T_{\text{SSET}} = 0.3$ mK, providing access to the quantum ground state of a resonator similar to that reported here.

Received 17 February; accepted 30 June 2006.

1. Caves, C. M. Quantum limits on noise in linear amplifiers. *Phys. Rev. D* **26**, 1817–1839 (1982).
2. Braginsky, V. B. & Khalili, F. Ya. *Quantum Measurement* (Cambridge Univ. Press, Cambridge, 1995).
3. Clerk, A. A. Quantum-limited position detection and amplification: A linear response perspective. *Phys. Rev. B* **70**, 245306 (2004).
4. Abbott, B. et al. Upper limits on gravitational wave bursts in LIGO’s second science run. *Phys. Rev. D* **72**, 062001 (2005).
5. Rugar, D., Budakian, R., Mamin, H. J. & Chui, B. W. Single spin detection by magnetic resonance force microscopy. *Nature* **430**, 329–332 (2004).
6. Titttonen, I. et al. Interferometric measurements of the position of a macroscopic body: Towards observations of quantum limits. *Phys. Rev. A* **59**, 1038–1044 (1999).

7. LaHaye, M. D., Buu, O., Camarota, B. & Schwab, K. C. Approaching the quantum limit of a nanomechanical resonator. *Science* **304**, 74–77 (2004).
8. Schoelkopf, R. J., Wahlgren, P., Kozhevnikov, A. A., Delsing, P. & Prober, D. The radio-frequency single-electron transistor (RF-SET): A fast and ultrasensitive electrometer. *Science* **280**, 1238–1242 (1998).
9. Fulton, T. A., Gammel, P. L., Bishop, D. J., Dunkleberger, L. N. & Dolan, G. J. Observation of combined Josephson and charging effects in small tunnel junction circuits. *Phys. Rev. Lett.* **63**, 1307–1310 (1989).
10. Clerk, A. A. & Bennett, S. Quantum nano-electromechanics with electrons, quasiparticles and Cooper pairs: effective bath descriptions and strong feedback effects. *N. J. Phys.* **7**, 238 (2005).
11. Blencowe, M. P., Imbers, J. & Armour, A. D. Dynamics of a nanomechanical resonator coupled to a superconducting single-electron transistor. *N. J. Phys.* **7**, 236 (2005).
12. Mozyrsky, D., Martin, I. & Hastings, M. B. Quantum-limited sensitivity of single-electron-transistor-based displacement detectors. *Phys. Rev. Lett.* **92**, 018303 (2004).
13. Armour, A. D., Blencowe, M. P. & Zhang, Y. Classical dynamics of a nanomechanical resonator coupled to a single-electron transistor. *Phys. Rev. B* **69**, 125313 (2004).
14. Mozyrsky, D. & Martin, I. Quantum-classical transition induced by electrical measurement. *Phys. Rev. Lett.* **89**, 018301 (2002).
15. Ojha, R. P., Lemieux, P.-A., Dixon, P. K., Liu, A. J. & Durian, D. J. Statistical mechanics of a gas-fluidized particle. *Nature* **427**, 521–523 (2004).
16. Clerk, A. A., Girvin, S. M., Nguyen, A. K. & Stone, A. D. Resonant Cooper-pair tunneling: quantum noise and measurement characteristics. *Phys. Rev. Lett.* **89**, 176804 (2002).
17. Gavish, U., Levinson, Y. & Imry, Y. Detection of quantum noise. *Phys. Rev. B* **62**, R10637 (2000).
18. Schoelkopf, R. J., Clerk, A. A., Girvin, S. M., Lehnert, K. W. & Devoret, M. H. in *Quantum Noise in Mesoscopic Physics* (Kluwer Academic, Norwell, Massachusetts, 2003).
19. Deblock, R., Onac, E., Gurevich, L. & Kouwenhoven, L. P. Detection of quantum noise from an electrically driven two-level system. *Science* **301**, 203–206 (2003).
20. Braginsky, V. B. & Vyatchanin, S. P. Low quantum noise tranquilizer for Fabry-Perot interferometer. *Phys. Lett. A* **293**, 228–234 (2002).
21. Lett, P. D. *et al.* Optical molasses. *J. Opt. Soc. Am. B* **6**, 2084–2107 (1989).
22. Pohlen, S. L., Fitzgerald, R. J. & Tinkham, M. The Josephson-quasiparticle (JQP) current cycle in the superconducting single-electron transistor. *Physica B* **284–288**, 1812–1813 (2000).
23. Choi, M.-S., Plastina, F. & Fazio, R. Charge and current fluctuations in a superconducting single-electron transistor near a Cooper pair resonance. *Phys. Rev. B* **67**, 045105 (2003).
24. Ruskov, R., Schwab, K. & Korotkov, A. N. Squeezing of a nanomechanical resonator by quantum nondemolition measurement and feedback. *Phys. Rev. B* **71**, 235407 (2005).
25. Armour, A. D., Blencowe, M. P. & Schwab, K. C. Entanglement and decoherence of a micromechanical resonator via coupling to a Cooper-pair box. *Phys. Rev. Lett.* **88**, 148301 (2002).
26. Irish, E. K. & Schwab, K. Quantum measurement of a coupled nanomechanical resonator-Cooper-pair box system. *Phys. Rev. B* **68**, 155311 (2003).
27. Martin, I., Shnirman, A., Tian, L. & Zoller, P. Ground-state cooling of mechanical resonators. *Phys. Rev. B* **69**, 125339 (2004).
28. Wilson-Rae, I., Zoller, P. & Imamoglu, A. Laser cooling of a nanomechanical resonator mode to its quantum ground state. *Phys. Rev. Lett.* **92**, 075507 (2004).
29. Lundin, U. Localized heating or cooling in mesoscopic devices by electron transport. *Phys. Lett. A* **332**, 127–130 (2004).
30. Nakamura, Y., Pashkin, Yu. A. & Tsai, J. S. Coherent control of macroscopic quantum states in a single-Cooper-pair box. *Nature* **398**, 786–788 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Rimberg and A. Vandaley for discussions, and B. Camarota for assistance with the fabrication of the samples. M.P.B. is supported by the NSF through an NIRT grant, A.D.A. is supported by the EPSRC, and A.A.C. is supported by NSERC.

Author Contributions A.N. and O.B. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.C.S. (schwab@ccmr.cornell.edu).

Oscillatory dependence of current-driven magnetic domain wall motion on current pulse length

Luc Thomas¹, Masamitsu Hayashi^{1,2}, Xin Jiang¹, Rai Moriya¹, Charles Rettner¹ & Stuart S. P. Parkin¹

Magnetic domain walls, in which the magnetization direction varies continuously from one direction to another, have long been objects of considerable interest¹. New concepts for devices based on such domain walls are made possible by the direct manipulation of the walls using spin-polarized electrical current^{2,3} through the phenomenon of spin momentum transfer^{4,5}. Most experiments to date have considered the current-driven motion of domain walls under quasi-static conditions^{6–12}, whereas for technological applications, the walls must be moved on much shorter timescales. Here we show that the motion of domain walls under nanosecond-long current pulses is surprisingly sensitive to the pulse length. In particular, we find that the probability of dislodging a domain wall, confined to a pinning site in a permalloy nanowire, oscillates with the length of the current pulse, with a period of just a few nanoseconds. Using an analytical model^{13–17} and micromagnetic simulations, we show that this behaviour is connected to a current-induced oscillatory motion of the domain wall. The period is determined by the wall's mass¹⁸ and the slope of the confining potential. When the current is turned off during phases of the domain wall motion when it has enough momentum, the domain wall is driven out of the confining potential in the opposite direction to the flow of spin angular momentum. This dynamic amplification effect could be exploited in magnetic nanodevices based on domain wall motion.

Most studies of current-induced motion of magnetic domain walls (DWs) have been carried out in magnetic nanowires formed from permalloy (Ni₈₁Fe₁₉)^{6–10}, one of the more attractive magnetic materials for nanodevices because of its low magnetic anisotropy and magnetization yet high Curie temperature¹⁹. Moreover, the electrical current in permalloy is highly spin polarized owing to significant spin-dependent scattering¹⁹. Experiments using direct currents or long pulses have shown that in permalloy nanowires, DWs move in the direction of the flow of the electrons (that is, opposite to the electrical current direction) when the current density exceeds a threshold value. In the present work we find that DWs can be moved with nanosecond-long current pulses smaller than the d.c. threshold, but that, under these circumstances, the probability that the DW moves oscillates with the length of the current pulse with a period of a few nanoseconds. Moreover, the DW moves in the opposite direction to the flow of spin-polarized electrons.

We study magnetic domain walls in permalloy nanowires, 200 nm wide and 40 nm thick. The wires are composed of two straight sections oriented at 90° to one another, connected by a curved portion with a radius of curvature of ~3 μm (see Fig. 1a). A single DW is formed in the curved region of the nanowire by applying an appropriate sequence of magnetic fields, as described in Methods. The DW position and its form, whether separating domains with magnetization directions pointing away (tail-to-tail, TT) or towards

(head-to-head, HH) one another, depends upon this sequence of fields. Magnetic force microscopy reveals that the DW has a vortex structure (Fig. 1b), as expected from micromagnetic simulations for the size and properties of the wire used^{20,21}.

Three electrical connections are made to the nanowire—labelled c_i ($i = 1, 2, 3$) in Fig. 1a—which are used both to inject current pulses and to measure the resistance of portions of the wire, R_{ij} , between contacts i and j (see Methods). A DW in the nanowire is readily detected, through the phenomenon of anisotropic magnetoresistance, from a slight drop in the wire's resistance due to the transverse component of the DW's magnetization. As illustrated in Fig. 1c, a small drop in wire resistance between the contacts c_1 and c_2 , $\Delta R_{12} \approx 0.15 \Omega$, is observed when a DW is formed between these contacts. The DW can be moved along the wire by a field, and its motion from the lower to the upper side of the contact c_2 monitored from a small increase in R_{12} (Fig. 1c) (and a corresponding decrease

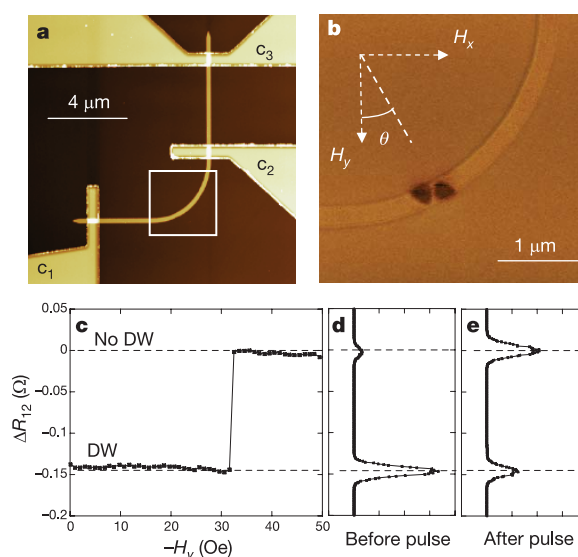


Figure 1 | Experimental configuration. **a**, Atomic force microscopy image of the device; c_1 , c_2 and c_3 are contacts. **b**, Magnetic force microscopy image corresponding to the area enclosed by the white square in **a**. A vortex DW is located in the bend of the wire at a position defined by the angle $\theta \approx 30^\circ$, by using a sequence of in-plane magnetic fields along the two perpendicular directions H_x and H_y . **c**, The change in the wire's resistance between contacts c_1 and c_2 , ΔR_{12} , as the HH DW is moved by field $-H_y$ along $-y$ across the contact c_2 without current. Dashed lines depict the wire's resistance with and without a DW. **d**, **e**, Histograms of the resistance levels measured at $H_y = -24$ Oe before (**d**) and after (**e**) the application of a current pulse ($V_p = -2.0$ V, $0.5 < t_p < 30$ ns).

¹IBM Almaden Research Center, 650 Harry Road, San José, California 95120, USA. ²Department of Materials Science and Engineering, Stanford University, Stanford, California 94305, USA.

in R_{23} , not shown). At zero current, the magnetic field along y (H_y) required to move the DW from its initial position across c_2 is $\sim \pm 36$ Oe for TT and HH DWs, respectively. The DW motion occurs when the component of H_y tangential to the wire at the DW position overcomes local pinning (~ 25 Oe), which probably results from edge roughness in the curved portion of the wire.

Current-driven DW motion is explored by first creating a DW in the nanowire and then by injecting a voltage pulse into the wire between contacts c_1 and c_2 . Measurements of ΔR_{12} before and after the pulse reveal whether the DW has moved across the contact c_2 (Fig. 1d and e). The probability of DW motion, P_M , is determined as a function of the magnetic field H_y , the voltage pulse amplitude V_p and the pulse length t_p by repeating the same sequence of DW injection and current pulsing 30 times for each set of parameters. Figure 2a–d show P_M for both HH and TT DWs as a function of t_p and H_y , for a constant pulse voltage of ± 2.0 V. Both types of DWs exhibit similar behaviours, whereas the dependence of P_M on the pulse length is strongly asymmetric with voltage polarity. DW motion occurs only with a very low probability below a threshold field (for the currents considered in these experiments). For positive voltages (electrons flowing from c_1 to c_2), the threshold field, which varies little with pulse length for pulses longer than ~ 3 ns, is ~ 10 Oe and ~ -5 Oe for TT and HH DWs, respectively. Above this threshold, P_M increases rapidly to $\sim 100\%$. On the contrary, for negative voltages, pronounced periodic oscillations of the DW motion probability are observed as the current pulse length is increased for fields ranging from 5 Oe to 30 Oe. The period of the oscillations is ~ 3 ns and ~ 4 ns for TT and HH DWs, respectively, but the number of oscillations observed increases with increasing field until the field is so large that

the DWs are always depinned, independent of the current pulse length.

The dependence of P_M on the pulse amplitude and length is shown at a constant field $H_y \approx 24$ Oe for TT DWs in Fig. 2e and f. Again, we observe a strong asymmetry with the pulse polarity. For positive voltages, DW motion occurs above a voltage threshold (~ 1 V) that is insensitive to the pulse length. For negative voltages, an oscillatory behaviour is observed. More than 10 oscillations can clearly be identified. The period of these oscillations increases for higher voltages, from ~ 2.9 ns for $V_p = 1.5$ V, up to ~ 3.9 ns for 3.2 V. We find that the oscillatory DW depinning is a general phenomenon, which we have observed not only in other permalloy nanowires of various widths (~ 100 – 300 nm) and thicknesses (~ 10 – 40 nm), but also in other materials (for example, FeCoNi) and structures (for example, spin-valve nanowires).

Our results appear consistent with the spin-transfer torque mechanism, in which the transfer of spin angular momentum from a polarized electrical current to the magnetization of the nanowire can induce DW motion^{2,3,13–17}, although the oscillatory behaviour we observe corresponds to DWs actually moving against the electron flow. Note that any field-related mechanism induced by the current pulse should give rise to opposite voltage polarity dependences for HH and TT DWs, which we do not observe.

We now develop an understanding of the oscillatory DW depinning using the well-established one-dimensional (1D) model¹. In this model, the profile of the DW is assumed to be unchanged during its motion so that its dynamics can be described by just two parameters, its position and its conjugate momentum. The latter is defined as $2M_S\Psi/\gamma$, where Ψ is the tilt angle of the wall magnetization out of the plane of the wire (see Fig. 3g), M_S is the saturation magnetization (~ 800 e.m.u. cm^{-3} for permalloy) and γ is the gyromagnetic factor (17.6 MHz Oe $^{-1}$). The 1D model has recently been generalized to include a spin-torque term^{13–17}, so that the Landau–Lifshitz–Gilbert equations of motion become:

$$(1 + \alpha^2)\dot{q} = -\frac{\alpha\gamma\Delta}{2M_S}\left(\frac{\partial\epsilon}{\partial q}\right) + \frac{\gamma\Delta}{2}H_k\sin(2\Psi) + (1 + \alpha\beta)u$$

$$(1 + \alpha^2)\dot{\Psi} = -\frac{\gamma}{2M_S}\left(\frac{\partial\epsilon}{\partial q}\right) - \frac{\gamma\alpha}{2}H_k\sin(2\Psi) + (\beta - \alpha)\frac{u}{\Delta}$$

where q and Δ are the DW's position and width, respectively, α is the Gilbert damping parameter, and H_k is the shape anisotropy field in the plane perpendicular to the wire's length that keeps the magnetization in the plane of the wire. The spin torque term is $u = \mu_B J P / e M_S$, with e the electron charge, J the current density, P the polarization of the current, and μ_B the Bohr magneton. When u is positive, the DW is driven in the direction of increasing q . β arises from a non-adiabatic contribution to the spin transfer torque^{14–17}. The term $\epsilon(q)$ is the DW potential energy per unit cross-sectional area, which includes a magnetic field along the DW propagation direction and a pinning potential term. The DW is pinned at a defect in the nanowire. For simplicity, we assume a parabolic potential of depth V and spatial extension q_0 such that, at $H = 0$, $\epsilon(q) = Vq_0(q/q_0)^2$ for $|q| < q_0$ and $\epsilon(q) = Vq_0$ for $|q| \geq q_0$.

Let us consider first the simplest case where $\beta = 0$ and $H = 0$. When a current smaller than the d.c. critical depinning current is applied, both q and Ψ oscillate in time (Fig. 3a, b) towards the equilibrium position $(0, \Psi_{\text{eq}}(u))$, with an oscillation frequency ω . The DW trajectory in the phase space (q, Ψ) is thus a spiral for which the oscillations are damped according to α (Fig. 3c). However, if the current is cut off before the DW reaches the final equilibrium position, the trajectory of the DW switches to a different orbit centred around the new equilibrium position without current, $(0, 0)$, as illustrated in Fig. 3d–f. Depending on the DW's momentum (that is, magnitude of Ψ) when the current is cut off, the amplitude of the oscillations is amplified (red curves in Fig. 3d and f), and the DW can leave the pinning potential. This is remarkable, as the DW is

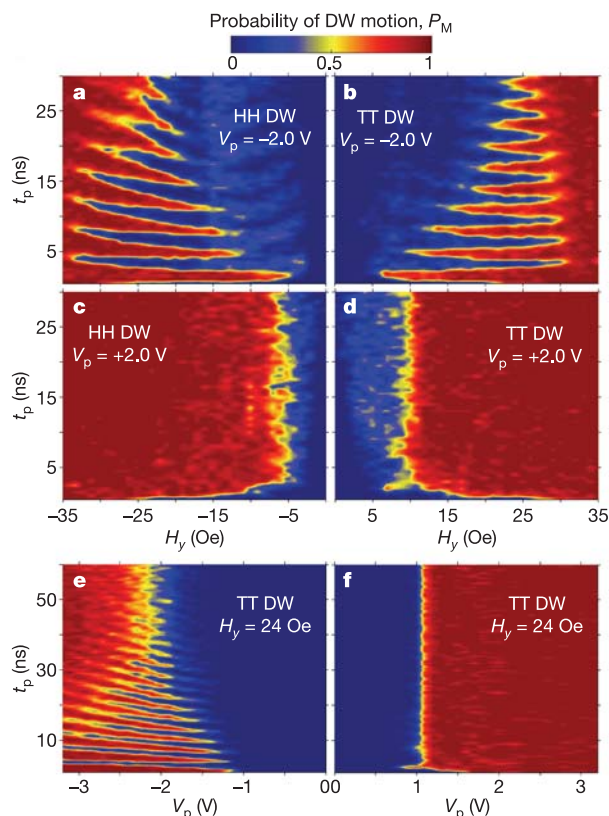


Figure 2 | Probability of motion of a domain wall subjected to current pulses of various lengths and amplitudes. Contour plots of the probability of current-driven DW motion P_M (colour scale) as a function of the current pulse length t_p , the pulse amplitude V_p and the magnetic field H_y ($H_x \approx 0$). Data points are taken every 1 Oe, 0.5 ns and 100 mV. **a–d**, Data for varying t_p and H_y at indicated values of V_p for HH (**a**, **c**) and TT (**b**, **d**) DWs; **e**, **f**, Data for varying t_p and V_p at indicated values of H_y . HH and TT stand for head-to-head and tail-to-tail DWs, respectively.

otherwise pinned for intermediate or longer current pulses. Moreover, the DW leaves in the opposite direction to that when driven by a current larger than the d.c. threshold current. Red lines in Fig. 3a–c show the current pulse cut-off times for which the DW leaves the pinning potential. The oscillation frequency is calculated from the equations of motion using a linear approximation ($\Psi \ll 1$): $\omega \approx \gamma[VH_k\Delta/(M_S q_0)]^{1/2}$. Note that this is simply the resonance frequency of a free harmonic oscillator of mass $m = 2M_S/(\gamma^2 H_k \Delta)$ (that is, the DW mass) in a restoring force $-2V/q_0$ (that is, the slope of the pinning potential). Both the DW mass and the restoring force are normalized per unit cross-sectional area.

In order to validate the results of the 1D model, we have carried out micromagnetic simulations²² of nanowires with different DW structures, both transverse (Fig. 3g) and vortex (Fig. 3h). Pinning of the DW was mimicked by including a small non-magnetic defect at one edge of the wire (see Methods). The main features found using the 1D

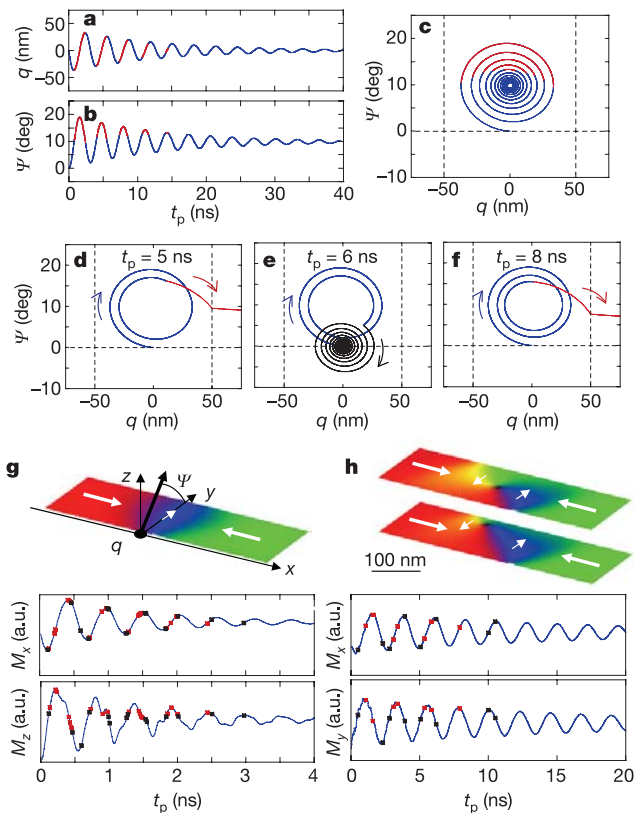


Figure 3 | Current-driven dynamics of a domain wall pinned in a parabolic potential well calculated with the 1D model. The Gilbert damping is $\alpha = 0.005$, the non-adiabatic contribution is $\beta = 0$, and the current parameter $u = -80 \text{ m s}^{-1}$; other parameters are described in Methods. Note that thermal effects are not included. **a–c**, Time dependence of the DW position q (**a**) and momentum Ψ (**b**) and the DW trajectory in the phase space (q, Ψ) for direct current (**c**). Red lines show the pulse length for which the DW leaves the potential well after the pulse. Vertical dashed lines indicate the width of the potential well ($\pm 50 \text{ nm}$). **d–f**, Examples of the DW trajectory in phase space for different pulse lengths. **g, h**, The bottom panels show colour maps of micromagnetic simulations (at 0 K) of permalloy nanowires with transverse (**g**) and vortex (**h**) DW structures. For the vortex wall (**h**), colour maps are shown at rest (top) and during the gyrotropic motion of the vortex core (bottom). Both nanowires are 100 nm wide, and their thickness is 5 nm (**g**) and 20 nm (**h**). The adiabatic spin transfer torque and the Oersted self-field are included in the simulations ($\beta = 0$). Gilbert damping is $\alpha = 0.01$. Solid lines show the net magnetization components M_x ($\sim$ DW position), and M_y or M_z ($\sim$ DW momentum, see text) for a current of 3 mA (current densities of 6.0×10^8 and $1.5 \times 10^8 \text{ A cm}^{-2}$, respectively) and a magnetic field $H_x = 10 \text{ Oe}$. Red and black symbols show whether the DW leaves or remains trapped at the notch after the current is turned off for different pulse lengths.

model are reproduced for both DW structures. For currents smaller than the critical current, the DW centre oscillates within the potential well around its equilibrium position. The comparison with the 1D model is straightforward for the transverse wall (Fig. 3g). The oscillation in position ($\sim M_x$) is associated with a slight tilt of the magnetization out of the plane of the wire ($\sim M_z$), as described by the conjugate momentum Ψ in the 1D approximation. For DWs with vortex structures (Fig. 3h), the oscillation of the DW within the potential well is associated with a gyrotropic motion of the vortex core, quite similar to the field^{23,24} or current driven dynamics²⁵ of a vortex core confined in a sub-micrometre disk. Magnetization maps of the vortex wall at rest and when distorted during its motion are shown in Fig. 3h. The net magnetization perpendicular to the wire's long axis ($\sim M_y$) oscillates, providing a good measure of the conjugate momentum Ψ . For both DW structures, the DW leaves the potential well only if its momentum is large enough when the current is cut off. Red and black symbols in the lower panels of Fig. 3g and h illustrate the pulse lengths for which the DW is or is not depinned, respectively. Note that the period of the oscillations increases with the dimensions of the nanowire from 0.5 ns (width 100 nm, thickness 5 nm; Fig. 3g) to 3 ns (width 200 nm, thickness 40 nm; not shown) and is in good agreement with the analytical expression calculated from the 1D model.

Micromagnetic simulations exhibit complex dynamical processes beyond the reach of the 1D approximation. For example, the DW structure can be distorted during the current pulse, and thus influence the DW motion. Nevertheless, the 1D model appears to describe well the essential physical mechanisms responsible for the oscillatory dependence of the DW motion on the pulse duration. We now show that this model can be used to describe our experiments quantitatively. Realistic parameters are estimated from micromagnetic calculations (see Methods). Current-driven DW motion is mapped for a TT DW as a function of t_p and H_y at constant u (Fig. 4a and b), and as a function of t_p and u at constant field (Fig. 4c and d). These maps are in good agreement with the experiments. For positive currents (electron flow along the field-driven motion direction), the threshold for depinning the DW is roughly independent of the length of the pulse because the DW leaves the pinning potential during the pulse. On the contrary, for negative currents, depinning occurs after the current is cut off. By fitting the experimental data to this model, we can deduce the

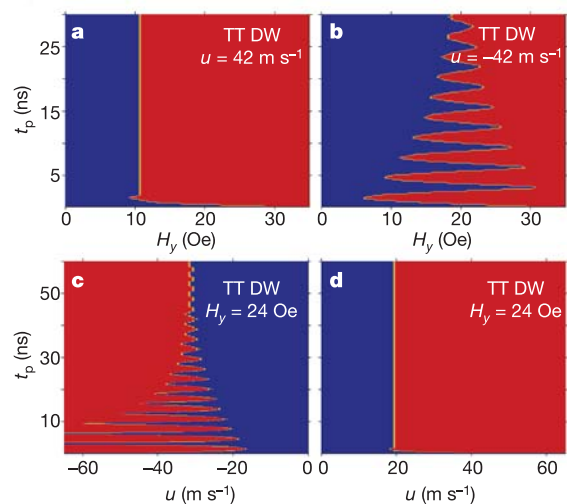


Figure 4 | Contour maps of the domain wall response to current pulses calculated within the 1D model. **a–d**, t_p is the pulse length, H_y the magnetic field, and u , the spin-torque parameter (proportional to current). The model parameters are deduced from micromagnetic simulations (see Methods). Regions in which DW motion does or does not occur are shown in red and blue, respectively.

magnitude of the damping constant, $\alpha \approx 0.005$, the non-adiabatic parameter, $\beta \approx 0.04$, and the current polarization, $P \approx 0.4$. The large value of β ($\sim 8\alpha$) suggests that the non-adiabatic spin transfer torque plays a significant role in the current-driven DW motion mechanism^{15,16}. Although our model does not account for certain details of our experimental results, such as the different oscillation periods for HH and TT DWs and the voltage dependence of these periods, the model clearly reveals the physical origin of the oscillatory depinning phenomenon.

We have observed an unusual phenomenon in which current-induced DW oscillations are amplified after the end of the current pulse, if the timing of the pulse matches the DW precession period. Taking advantage of this dynamic amplification could open ways of reducing the critical current density needed to move DWs in magnetic nanodevices.

METHODS

Experimental set-up. The nanowire was fabricated by electron-beam lithography from a 40-nm-thick permalloy film grown by magnetron sputtering deposition on a high-resistivity Si wafer. The film was capped with 1 nm TaN/5 nm Ru to prevent oxidation. The nanowire was coated on the sides with Al_2O_3 to limit oxidation along its edges. Another lithographic step was used to fabricate three contact pads (45-nm-thick Rh), labelled c_1 , c_2 and c_3 in Fig. 1a. A single DW is formed in the nanowire by first applying a magnetic field of ~ 700 Oe, larger than the shape anisotropy of the wire, to align the wire's magnetization along a direction tilted at an angle $\theta \approx 85^\circ$ to the wire (Fig. 1b). When this field is reduced to zero, the magnetizations in the wire's two straight sections orient along their respective lengths and a DW is formed in the curved region of the nanowire. In a second step, the DW is moved down and around the bend by applying a field (~ 80 Oe) along the y direction ($\theta = 0^\circ$). The form of the DW, whether tail to tail or head to head, depends on whether negative or positive fields are used. The initial position of the DW is then at $\theta \approx 44^\circ$.

Voltage pulses, with rise and fall times of about 0.1 ns, are injected in the nanowire between contacts c_1 and c_2 using a 7 GHz bandwidth probe. The d.c. magnetoresistance of the wire is measured using the same probe through a bias tee. As the change in resistance associated with the DW (ΔR_{12}) is small, the device resistance is compared to a reference state without a DW by making a measurement in zero field after saturating the wire in 700 Oe along $\theta = -60^\circ$. A small d.c. test current (~ 0.25 mA) is applied to the sample throughout the measurements. The conversion between the nominal voltage pulse V_p and the actual current pulse I flowing through the device is $I = 2V/(R + Z_0)$, where $R \approx 310 \Omega$ is the device d.c. resistance between c_1 and c_2 at room temperature (including the contribution of the two contact pads, $\sim 80 \Omega$) and $Z_0 = 50 \Omega$ is the characteristic impedance of the transmission line. The current through the device is $\sim 5.5 \text{ mA V}^{-1}$, corresponding to a current density of $\sim 6.9 \times 10^7 \text{ A cm}^{-2}$. For voltages higher than ~ 1 V, Joule heating during the pulse causes the device resistance to increase. Time-resolved measurements performed using a sampling oscilloscope show that the device resistance increases by a factor up to 1.3 for the highest pulse voltage used in this study (3.2 V), corresponding to a temperature rise of ~ 100 K. As the device resistance and the current vary with the pulse length, the pulse amplitude is defined using the nominal output voltage V_p rather than the current.

Determination of the parameters for the 1D model. Realistic values of the DW width parameter Δ and the transverse anisotropy H_k are estimated from micromagnetic simulations of a nanowire (200 nm wide, 40 nm thick) with a vortex DW. We find the dynamical DW width $\Delta \approx 15$ nm, in good agreement with ref. 21. Note that the discrepancy between this very small value and the apparent physical size of the DW (Fig. 1b) results from the strong influence of the vortex core on the DW motion. The transverse anisotropy $H_k = 1,800$ Oe is estimated from the Walker breakdown field, above which the DW velocity decreases strongly¹. The width of the pinning potential, $q_0 = 50$ nm, is obtained by simulating the quasi-static field-driven depinning of the DW from a small non-magnetic defect (10 nm long, 5 nm wide) introduced at one edge of the wire. Note that q_0 is not only determined by the physical size of the pinning defect but also by the DW profile. The magnitude of the pinning potential $V = 2 \times 10^4 \text{ erg cm}^{-3}$ is calculated from the formula derived in the text for the oscillation period so as to match the experimental value of ~ 3 ns.

For the calculations presented in Fig. 4, the pressure on the DW, $\partial \varepsilon(q)/\partial q$, depends on the pinning potential and the projection of the magnetic field H_y

along the bend:

$$\partial \varepsilon(q)/\partial q = 0 \text{ for } q < 0 \text{ (horizontal arm of the wire)}$$

$$\partial \varepsilon(q)/\partial q = -2H_y M_s \sin(q/R_c) \text{ for } 0 < q < \pi R_c/2 \text{ and } |q - q_i| > q_0$$

$$\partial \varepsilon(q)/\partial q = -2H_y M_s \sin(q/R_c) + 2V(q - q_i)/q_0 \text{ for } 0 < q < \pi R_c/2 \text{ and } |q - q_i| < q_0$$

$$\partial \varepsilon(q)/\partial q = -2H_y M_s \text{ for } q > \pi R_c/2 \text{ (vertical arm of the wire)}$$

where $R_c = 3 \mu\text{m}$ is the radius of curvature of the bend, and q_i is the centre of the pinning potential. The experimental value for the depinning field, $H_y \approx 36$ Oe, is obtained for $q_i = 2.3 \mu\text{m}$ (corresponding to $\theta \approx 44^\circ$). In our calculations, the DW is at its equilibrium position in the field when the current is turned on. We assume a rectangular pulse shape (zero rise and fall times).

Received 12 April; accepted 17 July 2006.

1. Malozemoff, A. P. & Slonczewski, J. C. *Magnetic Domain Walls in Bubble Material* (Academic, New York, 1979).
2. Berger, L. Exchange interaction between ferromagnetic domain wall and electric current in very thin metallic films. *J. Appl. Phys.* **55**, 1954–1956 (1984).
3. Berger, L. Possible existence of a Josephson effect in ferromagnets. *Phys. Rev. B* **33**, 1572–1578 (1986).
4. Slonczewski, J. C. Current-driven excitation of magnetic multilayers. *J. Magn. Mater.* **159**, L1–L7 (1996).
5. Berger, L. Emission of spin waves by a magnetic multilayer traversed by a current. *Phys. Rev. B* **54**, 9353–9358 (1996).
6. Grollier, J. *et al.* Switching a spin valve back and forth by current-induced domain wall motion. *Appl. Phys. Lett.* **83**, 509–511 (2003).
7. Vernier, N., Allwood, D. A., Atkinson, D., Cooke, M. D. & Cowburn, R. P. Domain wall propagation in magnetic nanowires by spin-polarized current injection. *Europhys. Lett.* **65**, 526–532 (2004).
8. Yamaguchi, A. *et al.* Real-space observation of current-driven domain wall motion in submicron magnetic wires. *Phys. Rev. Lett.* **92**, 077205 (2004).
9. Klau, M. *et al.* Controlled and reproducible domain wall displacement by current pulses injected into ferromagnetic ring structures. *Phys. Rev. Lett.* **94**, 106601 (2005).
10. Klau, M. *et al.* Direct observation of domain-wall configurations transformed by spin currents. *Phys. Rev. Lett.* **95**, 026601 (2005).
11. Yamanouchi, M., Chiba, D., Matsukura, F. & Ohno, H. Current-induced domain-wall switching in a ferromagnetic semiconductor structure. *Nature* **428**, 539–542 (2004).
12. Ravelosona, D., Lacour, D., Katine, J. A., Terris, B. D. & Chappert, C. Nanometer scale observation of high efficiency thermally assisted current-driven domain wall depinning. *Phys. Rev. Lett.* **95**, 117203 (2005).
13. Li, Z. & Zhang, S. Domain-wall dynamics driven by adiabatic spin-transfer torques. *Phys. Rev. B* **70**, 024417 (2004).
14. Tataru, G. & Kohno, H. Theory of current-driven domain wall motion: Spin transfer versus momentum transfer. *Phys. Rev. Lett.* **92**, 086601 (2004).
15. Thiaville, A., Nakatani, Y., Miltat, J. & Suzuki, Y. Micromagnetic understanding of current-driven domain wall motion in patterned nanowires. *Europhys. Lett.* **69**, 990–996 (2005).
16. Zhang, S. & Li, Z. Roles of nonequilibrium conduction electrons on the magnetization dynamics of ferromagnets. *Phys. Rev. Lett.* **93**, 127204 (2004).
17. Barnes, S. E. & Maekawa, S. Current-spin coupling for ferromagnetic domain walls in fine wires. *Phys. Rev. Lett.* **95**, 107204 (2005).
18. Saitoh, E., Miyajima, H., Yamaoka, T. & Tataru, G. Current-induced resonance and mass determination of a single magnetic domain wall. *Nature* **432**, 203–206 (2004).
19. O'Handley, R. C. *Modern Magnetic Materials: Principles and Applications* (Wiley and Sons, New York, 2000).
20. McMichael, R. D. & Donahue, M. J. Head to head domain wall structures in thin magnetic strips. *IEEE Trans. Magn.* **33**, 4167–4169 (1997).
21. Nakatani, Y., Thiaville, A. & Miltat, J. Head-to-head domain walls in soft nanostrips: a refined phase diagram. *J. Magn. Mater.* **290**, 750–753 (2005).
22. LLG Micromagnetic Simulator (<http://llgmicromagnetic.com/>) (1997).
23. Park, J. P., Eames, P., Engebretson, D. M., Berezovsky, J. & Crowell, P. A. Imaging of spin dynamics in closure domain and vortex structures. *Phys. Rev. B* **67**, 020403 (2003).
24. Novosad, V. *et al.* Magnetic vortex resonance in patterned ferromagnetic dots. *Phys. Rev. B* **72**, 024455 (2005).
25. Shibata, J., Nakatani, Y., Tataru, G., Kohno, H. & Otani, Y. Current-induced magnetic vortex motion by spin-transfer torque. *Phys. Rev. B* **73**, 020403 (2006).

Acknowledgements We acknowledge financial support from DMEA.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.T. (luthom@us.ibm.com) or S.S.P.P. (parkin@almaden.ibm.com).

Observation of an O₈ molecular lattice in the ϵ phase of solid oxygen

Lars F. Lundegaard¹, Gunnar Weck², Malcolm I. McMahon¹, Serge Desgreniers³ & Paul Loubeyre²

Of the simple diatomic molecules, oxygen is the only one to carry a magnetic moment. This makes solid oxygen particularly interesting: it is considered a 'spin-controlled' crystal¹ that displays unusual magnetic order². At very high pressures, solid oxygen changes from an insulating to a metallic state³; at very low temperatures, it even transforms to a superconducting state⁴. Structural investigations of solid oxygen began in the 1920s and at present, six distinct crystallographic phases are established unambiguously¹. Of these, the ϵ phase of solid oxygen is particularly intriguing: it exhibits a dark-red colour, very strong infrared absorption, and a magnetic collapse¹. It is also stable over a very large pressure domain and has been the subject of numerous X-ray diffraction^{5–7}, spectroscopic^{8–11} and theoretical studies^{12–14}. But although ϵ -oxygen has been shown to have a monoclinic $C2/m$ symmetry^{5–7,15} and its infrared absorption behaviour attributed to the association of oxygen molecules into larger units^{9,14}, its exact structure remains unknown. Here we use single-crystal X-ray diffraction data collected between 13 and 18 GPa to determine the structure of ϵ -oxygen. We find that ϵ -oxygen is characterized by the association of four O₂ molecules into a rhombohedral molecular unit, held together by what are probably weak chemical bonds. This structure is consistent with existing spectroscopic data, and further validated by the observation of a newly predicted Raman stretching mode.

X-ray studies of solid oxygen began in the 1920s, and, at present, six solid crystallographic modifications are unambiguously established¹. The α , β and γ phases exist at ambient pressure and low temperature. Under pressure at $T = 295$ K, oxygen solidifies into the β phase at 5.4 GPa, then transforms into the 'orange' δ phase at 9.6 GPa, which in turn undergoes a strongly first-order transition to the 'red' ϵ phase at 10 GPa. Above 96 GPa, the ϵ phase transforms into the metallic ζ phase^{3,4}. The α , β , δ phases are layered structures, and the non-magnetic rhombohedral β phase is an optimal packing of diatomic molecules possessing a small quadrupole moment. A small deformation of this structure due to the magnetic interaction results in the monoclinic or orthorhombic structures of the magnetic α and δ phases, respectively¹. The ϵ phase was discovered¹⁶ in 1979 with the first loading of oxygen in a diamond anvil cell. Drastic changes occur at the δ – ϵ transition: a significant volume discontinuity ($\sim 7\%$), the appearance of a strong infrared absorption, new Raman and infrared modes, a dramatic colour change to dark red, and a magnetic collapse^{1,17}. Optical and X-ray powder-diffraction studies have suggested that the ϵ phase retains the layered nature of the α , β and δ phases, with the O₂ molecules remaining aligned parallel to each other. Johnson *et al.*⁵ proposed that the ϵ phase has monoclinic $C2/m$ symmetry with eight molecules in the unit cell, but although subsequent powder^{6,15} and single-crystal⁷ diffraction experiments confirmed the $C2/m$ space group, the positions of the

molecules remained elusive. To explain the infrared absorption of the ϵ phase, a molecular association into O₄ units has been proposed⁹, and density functional calculations suggest a larger association of molecules, with the formation of herringbone-type chains of O₂ molecules¹⁴. We demonstrate here that an association of molecules does indeed take place in the ϵ -phase, but into units containing four O₂ molecules (and hence eight atoms).

The growth of a single crystal of the ϵ phase directly from the melt is difficult, and the $C2/m$ space group was determined previously from a sample consisting of at least seven crystals⁵. However, the ability to grow high-quality single crystals of the ϵ phase in a helium hydrostatic pressure medium now makes it possible to perform detailed structural studies of oxygen at high pressure. The first synchrotron X-ray diffraction study of such a crystal was made with the energy dispersive technique⁷ and, although the $C2/m$ space group was confirmed, the number of peaks was insufficient to determine the structure. However, high-pressure single-crystal diffraction techniques using charge-coupled device (CCD) area detectors and synchrotron radiation are ideally suited to determining complex structures using such samples¹⁸, and we have therefore used this technique to determine the structure of ϵ -oxygen between 13.2 and 17.6 GPa.

Two different ϵ -oxygen crystals were studied. For crystal 1, data were collected at 17.5 GPa and these revealed that the alignment of the crystal relative to the pressure cell was unfavourable, resulting in the measurement of only 33 unique reflections. The crystal was therefore re-grown in a more favourable orientation that enabled the measurement of 55 unique reflections at 17.6 GPa. After collecting data at this pressure, the pressure was reduced and further data were collected from crystal 2 at 15.6 GPa and 13.2 GPa. Figure 1a illustrates the combination of techniques used: the micrograph shows crystal 2 of ϵ -oxygen in helium at 17.6 GPa, and the diffraction pattern comprises 13 individual CCD images from the data collected at 17.6 GPa.

Both crystals were found to be twinned, but with different twin laws, and diffraction data from only a single twin component was used in each case (see Supplementary Information). Indexing of the observed reflections confirmed the symmetry of the ϵ phase as C -centred monoclinic at all pressures, while the systematic absences showed the space group to be $C2$, Cm or $C2/m$, in perfect agreement with the previous single-crystal diffraction study⁵. The lattice parameters at 17.6 GPa were $a = 7.770(10)$ Å, $b = 5.501(5)$ Å, $c = 3.657(3)$ Å and $\beta = 116.4(1)^\circ$, where the estimated standard deviations are indicated in parentheses. The structure of the ϵ phase was solved in the highest-symmetry $C2/m$ spacegroup from the data collected from crystal 2 at 17.6 GPa, using the structure solution software SIR92 (see Methods), with trial refinements in the lower-symmetry $C2$ and Cm space groups showing no significant

¹SUPA, School of Physics and the Centre for Science at Extreme Conditions, The University of Edinburgh, Edinburgh, EH9 3JZ, UK. ²Département de Physique Théorique et Appliquée, Commissariat à l'Énergie Atomique, 91680 Bruyères-le-Châtel, France. ³Department of Physics, University of Ottawa, Ontario K1N 6N5, Canada.

improvement in fit. The solved structure contains three symmetry-inequivalent atoms: two (oxygen O1 and O2) occupy $4i$ positions of the $C2/m$ space group at $(x,0,z)$, and the third (O3) an $8j$ position at (x,y,z) . The refined structural coordinates of ϵ -oxygen from the two crystals at 17.5 and 17.6 GPa, along with the final R -factor, are given in Table 1 (see also Supplementary Information). When tested, previously proposed structures for the ϵ phase gave significantly poorer fits (see also Supplementary Information). The structure of the ϵ phase at 17.6 GPa, as viewed perpendicular to the a - b plane, is shown in Fig. 2. As in the high-pressure β ¹⁹ and δ ²⁰ phases, ϵ -oxygen comprises layers of collinear O_2 molecules, the axes of which are oriented perpendicular to the layers which lie in the a - b plane. The striking feature of the structure is the association of four O_2 molecules into rhomb-shaped $(O_2)_4$ molecular units, which are symmetry equivalent and centred on the lattice points at $(0,0,0)$ and $(0.5,0.5,0)$. Within these molecular units, each molecule has two neighbours at a distance of 2.18(1) Å at 17.6 GPa and with intermolecular angles of 96° and 84°. The $(O_2)_4$ units observed at 17.6 GPa are also present at both 15.6 and 13.2 GPa, and they are therefore formed at the δ -to- ϵ transition rather than through compression of the ϵ phase.

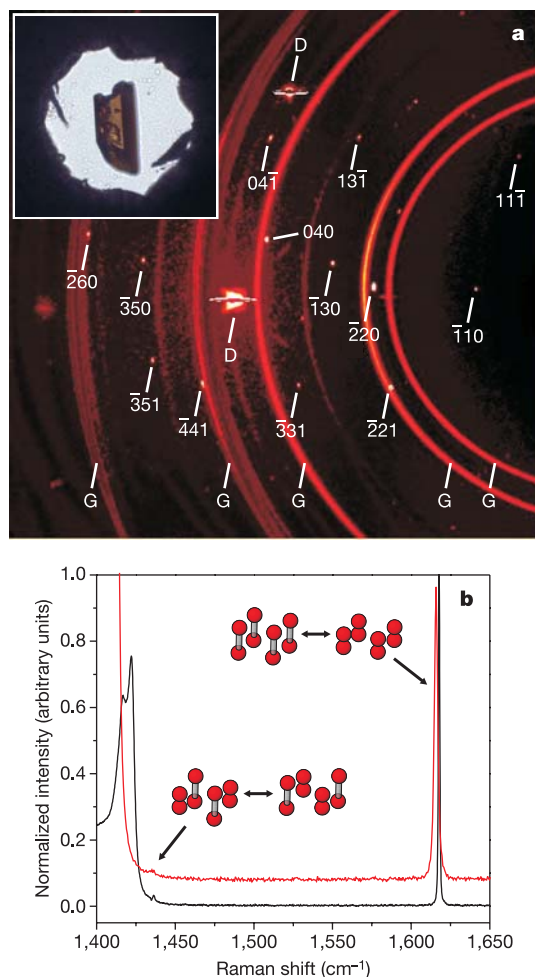


Figure 1 | X-ray and Raman measurements of ϵ -oxygen. **a**, Composite image of 13 individual CCD images showing the indexed Bragg peaks from the ϵ phase of oxygen. Diffraction features from the gasket and diamond anvils are marked G and D, respectively. The inset shows a micrograph of crystal 2 of ϵ -oxygen in helium at 17.6 GPa. **b**, Raman (red) and Fourier transform-Raman (black) spectra of two different crystals of ϵ -oxygen at 32.8 GPa and 33.4 GPa, respectively, showing the two stretching modes of the $(O_2)_4$ molecular unit. The motion of the atoms for the two modes is indicated.

The structural relationship between the β , δ and ϵ phases is illustrated in Fig. 3. The β and δ structures can both be described using the C -centred monoclinic cell of the α -phase, with the molecules remaining centred on the unit cell's lattice points^{1,19,20}. Any changes in the intermolecular distances within the molecular layers can then arise only via deformation of this unit cell. In contrast, the δ - ϵ transition involves a doubling of the δ -phase unit cell in both the a and b directions, and the relative displacement of half the molecules in each layer along the $\pm a$ direction (Fig. 3c). As a result, there is a dramatic reduction in the nearest-neighbour intermolecular distance from 2.57 Å in the δ phase at 9.6 GPa (ref. 20) to 2.25 Å in the ϵ phase at 13.2 GPa; this reduction gives rise to the large volume contraction observed at the δ - ϵ transition.

While molecular units containing eight oxygen atoms have not previously been observed in any oxygen phase, gas-phase $(O_2)_2$ dimers with an equilibrium spacing of 3.5 Å between the two parallel O_2 molecules are known^{21,22}. Pauling²³ proposed that the formation of such a dimer requires that the chemical spin-spin contribution should not be negligible, as confirmed recently²². The observed rhombohedral shape of the $(O_2)_4$ molecular units can probably also be explained by a bonding between open shell π^* orbitals, and $(O_2)_4$ and the gas-phase dimer $(O_2)_2$ should thus both be considered as chemically weakly bonded units. Indeed, while the O-O bond length within each O_2 molecule is unaltered by pressure, the inter- O_2 distances both within, and between, the $(O_2)_4$ groups compress at the same rate of 0.011(3)–0.014(3) Å GPa^{−1}, emphasizing the weakness of the interaction between the O_2 molecules that form the $(O_2)_4$ units. However, the intensity increase in the infrared modes of the ϵ phase up to 20–30 GPa has been taken as an indication that there is a progressive strengthening of the inter- O_2 bonds over this pressure range^{9,10}. Further detailed diffraction studies to higher pressures will be required to address this.

An $(O_2)_4$ complex with the observed D_{2h} symmetry has a large number of allowed vibrational modes: nine Raman and seven infrared fundamental modes. The vibrational modes were calculated for an isolated $(O_2)_4$ complex with the same structure as that determined at 17.6 GPa, using the Spartan 5.1 program (an *ab initio* quantum mechanics software package for computing molecular

Table 1 | Data collection and refined structural parameters

Pressure (GPa)	Crystal 1 17.5	Crystal 2 17.6
Unit cell		
Space group	$C2/m$	$C2/m$
a (Å)	7.760(10)	7.770(10)
b (Å)	5.510(3)	5.501(5)
c (Å)	3.656(8)	3.657(3)
β (°)	116.5(1)	116.4(1)
Data collection		
Number of unique reflections	33	55
Number of observed reflections (with $ F_{\text{obs}} > 4\sigma$)	25	46
$R_{\text{int}} = \sum [F_{\text{obs}}^2 - F_{\text{obs}}^2(\text{mean})] / \sum [F_{\text{obs}}^2]$	0.105	0.104
$R_{\sigma} = \sum [\sigma(F_{\text{obs}}^2)] / \sum [F_{\text{obs}}^2]$	0.064	0.044
Refinement		
R ($ F_{\text{obs}} > 4\sigma$)	0.069	0.038
R (for all)	0.083	0.062
Structure parameters		
O1 x	0.827(2)	0.826(1)
O1 z	0.173(5)	0.175(2)
O2 x	0.246(3)	0.247(1)
O2 z	0.190(6)	0.192(2)
O3 x	0.036(2)	0.0379(9)
O3 y	0.2673(5)	0.2666(6)
O3 z	0.180(4)	0.185(1)
O1 Uiso (Å ²)	0.013(2)	0.015(1)
O2 Uiso (Å ²)	0.014(2)	0.017(1)
O3 Uiso (Å ²)	0.013(1)	0.015(1)

Results are given for crystals 1 and 2 at 17.5 GPa and 17.6 GPa, respectively. The estimated standard deviations are indicated in parentheses. Details at other pressures are given in the Supplementary Information.

properties) with density functional theory²⁴. This simple calculation does not take into account effects such as the crystal field; but it nevertheless allows for a semiquantitative comparison with the spectroscopic data, yielding mode frequencies that agree within a few tens of cm^{-1} with the observed modes. Three infrared modes were found to have significant absorption (B_{3u} at $1,500\text{ cm}^{-1}$, B_{2u} at $1,443\text{ cm}^{-1}$ and B_{1u} at 292 cm^{-1}), in agreement with the reported infrared spectroscopic data^{9,11}. In addition, four A_g Raman modes should also be measurable, three of which, at $1,600\text{ cm}^{-1}$, 342 cm^{-1} and 161 cm^{-1} have already been observed⁸. The fourth Raman mode, at approximately $1,380\text{ cm}^{-1}$, is a signature of the $(\text{O}_2)_4$ unit but has not been observed to date, perhaps because it lies very close to the Raman signal from the diamond anvils of the pressure cell. This fourth mode is also expected to be very weak owing to the antisymmetric

stretching motion of the four O_2 molecules, coupled in diagonal pairs. Raman and Fourier-Transform Raman with excitation at 514.5 nm and $1,064\text{ nm}$, respectively, were carried out to search specifically for this peak. In different samples, a weak Raman-active line was observed to emerge from the Raman signal from the diamond anvils at pressures above 30 GPa . At 33 GPa , the new Raman line was observed at $1,430\text{ cm}^{-1}$, as illustrated in Fig. 1b. We interpret this new line to be the fourth Raman active mode of the $(\text{O}_2)_4$ molecular unit, and, given its weakness, it was best observed by performing Raman scattering on single-crystal samples with high-quality optical surfaces.

The rhomb-shaped $(\text{O}_2)_4$ molecular unit of the ϵ phase differs from the long-sought cyclooctaoxygen molecule, which would be the

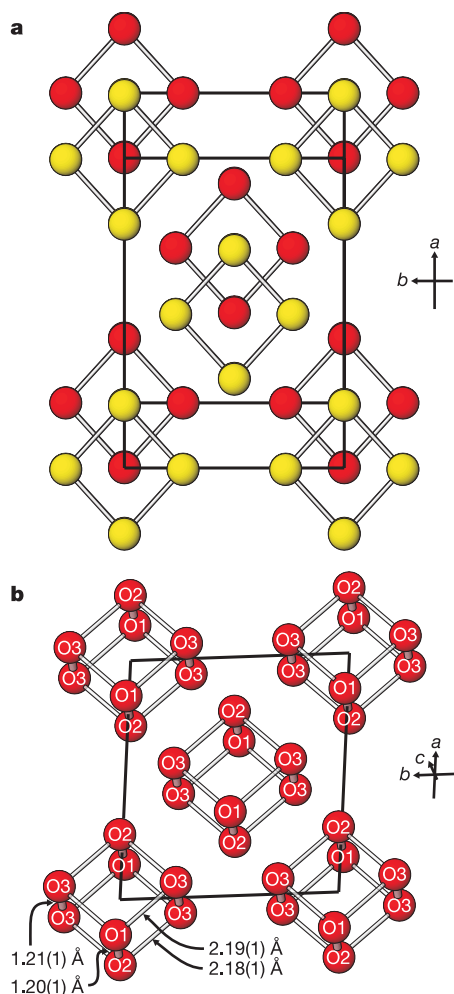


Figure 2 | The structure of ϵ -oxygen at 17.6 GPa . **a**, The structure as viewed perpendicular to the a - b plane. Red and yellow atoms lie in symmetry-equivalent molecular layers located at heights $z = 0$ and 1 , respectively. The groups of molecules forming the $(\text{O}_2)_4$ units are connected by thick lines. An inclined view of the $z = 0$ molecular layer is shown in **b**. The layers comprise two types of O_2 molecules—those containing two symmetry-equivalent O3 atoms and the centres of which are constrained by symmetry to lie exactly in the a - b planes at $z = 0$ and 1 , and those comprising symmetry-inequivalent O1 and O2 atoms, and the centres of which are unconstrained. The symmetry of the structure does not require that the O-O bond lengths of the O3-O3 and O1-O2 molecules be the same, nor that the molecules be oriented perpendicular to the a - b plane. However, the bond lengths of all the O_2 molecules are the same within error at all pressures, with a bond length of $1.20(3)\text{ Å}$ and in good agreement with the O-O bond length in the gas phase (1.207 Å)¹. Additionally, all the molecules are perpendicular to the a - b planes to within 1° .

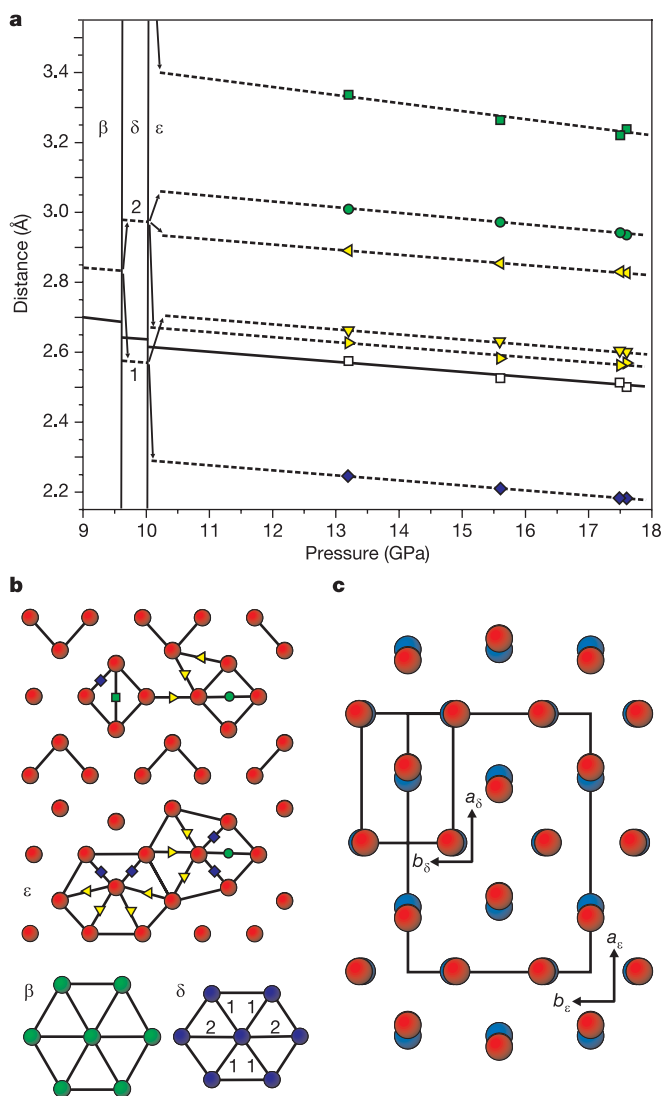


Figure 3 | Structural evolution of oxygen with pressure. **a**, Pressure dependence of the intermolecular distances within the molecular layers of the β , δ and ϵ phases of solid oxygen. The coloured symbols are identified in **b**. The shortest interlayer O-O distance in the ϵ phase is shown using unfilled squares, and the same distance is shown as a solid line for the β and δ phases. In the β phase, each molecule is surrounded within the layer by six others, all at the same distance¹⁹. The distortion within the molecular layer at the β - δ transition results in two different intermolecular distances in the δ phase²⁰, marked 1 and 2 in **a** and **b**. The larger unit cell and lower symmetry of the ϵ phase results in two symmetry-inequivalent coordination shells, and a wider range of intermolecular distances. **c**, The relationship between the locations of molecules within the molecular layers of the δ (blue atoms) and ϵ (red atoms) phases.

isoelectronic analogue of the S_8 rings in elemental sulphur²⁵. As may be inferred from the fact that the structure of the ϵ -phase of oxygen reported here has not been predicted by various *ab initio* calculations, it presents a challenge to our understanding of dense oxygen, particularly of the nature of the interactions that develop to stabilize the $(O_2)_4$ unit. We also note that the observation of molecular association in oxygen under pressure may affect explanations of the spectroscopic data from other elemental molecular solids (interestingly, the increased infrared absorption of phase III of solid hydrogen²⁶ parallels the case of the ϵ phase of oxygen), and the miscibility of O_2 with other elements at high pressure²⁷.

METHODS

Crystals of ϵ -oxygen were grown from a 2.6 mol% O_2 in He mixture at 450 K and 22.5 GPa, as described previously⁷. The diamond anvil pressure cell was equipped with anvil support seats with full conical apertures that enabled 70° angular X-ray access to the sample without contaminating scatter. X-ray diffraction data were collected on station 9.8 at the SRS, Daresbury Laboratory, using an incident X-ray wavelength of 0.688 Å. The pressure was measured using the standard ruby fluorescence technique. Diffraction data were measured over the full angular ranges permitted by the cell, and at two azimuthal angles ($\chi = 0^\circ$ and 90°) around the X-ray beam. No correction was made for the absorption of the sample and diamonds. Indexing and integration of the data were performed using the Bruker software suite²⁸. The structure was solved using SIR92²⁹, and refined using SHELX³⁰.

Received 2 May; accepted 16 August 2006.

- Freiman, Y. A. & Jodl, H. J. Solid oxygen. *Phys. Rep.* **401**, 1–228 (2004).
- Goncharenko, I. N., Makarova, O. L. & Ulivi, L. Direct determination of the magnetic structure of the delta phase of oxygen. *Phys. Rev. Lett.* **93**, 055502 (2004).
- Desgreniers, S., Vohra, Y. K. & Ruoff, A. L. Optical response of very high density solid oxygen to 132 GPa. *J. Phys. Chem.* **94**, 1117–1122 (1990).
- Shimizu, K., Suhara, K., Ikumo, M., Eremets, M. I. & Amaya, K. Superconductivity in oxygen. *Nature* **393**, 767–769 (1998).
- Johnson, S. W., Nicol, M. & Schiferl, D. Algorithm for sorting diffraction data from a sample consisting of several crystals enclosed in a sample environment apparatus. *J. Appl. Cryst.* **26**, 320–326 (1993).
- Desgreniers, S. & Brister, K. E. in *High Pressure Science and Technology* (ed. Trzeciakowski, W. A.) 363–365 (World Scientific, Singapore, 1996).
- Weck, G., Loubeyre, P. & LeToullec, R. Observation of structural transformations in metal oxygen. *Phys. Rev. Lett.* **88**, 035504 (2002).
- Akahama, Y. & Kawamura, H. High-pressure Raman spectroscopy of solid oxygen. *Phys. Rev. B* **54**, R15602–R15605 (1996).
- Gorelli, F. A., Ulivi, L., Santoro, M. & Bini, R. The ϵ phase of solid oxygen: Evidence of an O_4 molecule lattice. *Phys. Rev. Lett.* **83**, 4093–4096 (1999).
- Akahama, Y. & Kawamura, H. High-pressure infra-red spectroscopy of solid oxygen. *Phys. Rev. B* **61**, 8801–8805 (2000).
- Agnew, S. F., Swanson, B. I. & Jones, L. H. Extended interactions in the ϵ phase of oxygen. *J. Chem. Phys.* **86**, 5239–5245 (1987).
- Serra, S., Chiarotti, G., Scandolo, S. & Tosatti, E. Pressure-induced magnetic collapse and metallization of molecular oxygen: the ζ - O_2 phase. *Phys. Rev. Lett.* **80**, 5160–5163 (1998).
- Gebauer, R. *et al.* Noncollinear spin polarization from frustrated antiferromagnetism: A possible scenario for molecular oxygen at high pressure. *Phys. Rev. B* **61**, 6145–6149 (2000).
- Neaton, J. B. & Ashcroft, N. W. Low-energy linear-structures in dense oxygen: implications for the ϵ -phase. *Phys. Rev. Lett.* **88**, 205503 (2002).
- Akahama, Y., Kawamura, H., Häusermann, D., Hanfland, M. & Shimomura, O. New high-pressure structural transition of oxygen at 96 GPa associated with metallization in a molecular solid. *Phys. Rev. Lett.* **74**, 4690–4693 (1995).
- Nicol, M., Hirsch, K. R. & Holzapfel, W. B. Oxygen phase equilibria near 298 K. *Chem. Phys. Lett.* **68**, 49–52 (1979).
- Goncharenko, I. N. Evidence of magnetic collapse in the epsilon phase of solid oxygen. *Phys. Rev. Lett.* **94**, 205701 (2005).
- McMahon, M. I., Nemes, R. J. & Rekh, S. Complex crystal structure of cesium-III. *Phys. Rev. Lett.* **87**, 255502 (2001).
- Schiferl, D., Cromer, D. T. & Mills, R. L. Structure of O_2 at 5.5 GPa and 299 K. *Acta Crystallogr. B* **37**, 1329–1332 (1981).
- Schiferl, D., Cromer, D. T., Schwalbe, L. A. & Mills, R. L. Structure of 'orange' $^{18}O_2$ at 9.6 GPa and 297 K. *Acta Crystallogr. B* **39**, 153–157 (1983).
- Long, C. A. & Ewing, G. E. Spectroscopic investigation of van der Waals molecules. Infrared and visible spectra of $(O_2)_2$. *J. Chem. Phys.* **58**, 4824–4834 (1973).
- Aquilanti, V. *et al.* Quantum interference scattering of aligned molecules: binding in O_4 and role of spin coupling. *Phys. Rev. Lett.* **82**, 69–72 (1999).
- Pauling, L. *The Nature of the Chemical Bond* (Cornell Univ. Press, Ithaca, New York, 1964).
- SPARTAN version 5.1 (Wavefunction Inc., Irvine, California, 1998).
- Kim, K. S., Jang, J. H., Kim, S., Byung-Jin Min, B.-J. & Schaefer, H. F. Potential new high energy density materials: cyclooctaoxygen O_8 , including comparisons with the well-known cyclo- S_8 molecule. *J. Chem. Phys.* **92**, 1887–1892 (1990).
- Hanfland, M., Hemley, R. J. & Mao, H. K. Novel infrared vibron absorption in solid hydrogen at megabar pressures. *Phys. Rev. Lett.* **70**, 3760–3763 (1993).
- Sihachakr, D. & Loubeyre, P. O_2/N_2 mixtures under pressure: A structural study of the binary phase diagram at 295 K. *Phys. Rev. B* **70**, 134105 (2004).
- SMART, SAINT, ASTRO and XPREP: Data Collection and Processing Software for the SMART System (Bruker Analytical X-ray Systems, Inc., Madison, Wisconsin, 1995).
- Altomare, A., Casciaro, G., Giacovazzo, C. & Guagliardi, A. Completion and refinement of crystal structures with SIR92. *J. Appl. Cryst.* **26**, 343–350 (1993).
- Sheldrick, G. M. *SHELX97 Programs for Crystal Structure Analysis* (Release 97-2). (Univ. of Göttingen, 1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge discussions with T. Balic-Zunic on the sample twinning. We gratefully acknowledge the assistance of J. Warren and T. Prior in using beamline 9.8 at SRS, Daresbury Laboratory. The work was supported by research grants from the EPSRC, and facilities and other support from Daresbury Laboratory and the CCLRC. S.D. acknowledges the financial support of CEA/DAM Île-de-France and NSERC.

Author Contributions P.L. and G.W. prepared the ϵ -oxygen crystals. L.F.L. and M.I.M. performed the X-ray measurements and the structural analysis. G.W. performed the calculations of the modes. S.D. and G.W. performed the Raman measurements. P.L. and M.I.M. wrote most of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.I.M. (mim@ph.ed.ac.uk) or P.L. (paul.loubeyre@cea.fr).

Land-atmosphere coupling and climate change in Europe

Sonia I. Seneviratne¹, Daniel Lüthi¹, Michael Litschi¹ & Christoph Schär¹

Increasing greenhouse gas concentrations are expected to enhance the interannual variability of summer climate in Europe^{1–3} and other mid-latitude regions^{4,5}, potentially causing more frequent heatwaves^{1,3,5,6}. Climate models consistently predict an increase in the variability of summer temperatures in these areas, but the underlying mechanisms responsible for this increase remain uncertain. Here we explore these mechanisms using regional simulations of recent and future climatic conditions with and without land–atmosphere interactions. Our results indicate that the increase in summer temperature variability predicted in central and eastern Europe is mainly due to feedbacks between the land surface and the atmosphere. Furthermore, they suggest that land–atmosphere interactions increase climate variability in this region because climatic regimes in Europe shift northwards in response to increasing greenhouse gas concentrations, creating a new transitional climate zone with strong land–atmosphere coupling in central and eastern Europe. These findings emphasize the importance of soil-moisture–temperature feedbacks (in addition to soil-moisture–precipitation feedbacks^{7–10}) in influencing summer climate variability and the potential migration of climate zones with strong land–atmosphere coupling^{7,11} as a consequence of global warming. This highlights the crucial role of land–atmosphere interactions in future climate change.

Increases in climate variability have a greater effect on society than do changes in mean climate because it is more difficult to adapt to changes in extremes. Europe has experienced such extremes in recent years: the continent was struck by an unprecedented heatwave and serious drought in 2003 (refs 1, 12, 13), while cool summers with heavy precipitation and devastating floods occurred in 2002 (refs 14, 15) and 2005. In principle, these events are consistent with climate-change projections for this region: simulations driven by increasing greenhouse gas concentrations predict a considerable enhancement of interannual (year-to-year) variability of the European summer climate, both for temperature and precipitation, associated with higher risks of heatwaves^{1,5}, droughts¹⁶ and heavy precipitation events^{14,17}.

Here we focus on the projected increase in summer temperature variability in Europe, and on the underlying mechanisms responsible for it. Some studies have highlighted the role of interactions involving land-surface processes^{3,6} and changes in the radiation budget⁶, while others have emphasized future modifications in the summer atmospheric circulation^{5,15} and associated teleconnection patterns^{18,19}. In general, the latter hypothesis prevails, consistent with recent results from the Global Land–Atmosphere Coupling Experiment (GLACE)^{7,11}, suggesting that on average state-of-the-art atmospheric general circulation models (AGCMs) do not present any significant soil-moisture–precipitation or soil-moisture–temperature coupling in Europe. However, no similar analysis was performed to investigate possible changes of land–atmosphere coupling with increasing greenhouse gas concentrations.

The purpose of the present study was to isolate specifically the role of land–atmosphere coupling in projected changes in interannual climate variability during the extratropical summer season. The experiments (see Methods for details) consist of four 30-year-long climate experiments with a regional climate model (RCM), two for recent (CTL and CTL_{uncoupled}; 1960–1989) and two for future climatic conditions (SCEN and SCEN_{uncoupled}; 2070–2099). The CTL and SCEN experiments represent unperturbed simulations for the two time periods considered. The two additional simulations, CTL_{uncoupled} and SCEN_{uncoupled} share the same set-up as CTL and SCEN, except that their soil-moisture evolution at each time step is replaced with the climatology of CTL and SCEN, respectively. This removes the interannual variability of soil moisture and effectively uncouples the land surface from the atmosphere²⁰. To ensure that the simulated soil-moisture fields are in balance with the large-scale climate, the analysis is restricted to the last 20 years of the simulations.

We first analyse how the uncoupling of the land–atmosphere system affects summer temperature variability. Figure 1a–d displays the temperature variability in the four experiments, expressed in terms of the standard deviation of the summer (June–August) two-metre temperature T_{2m} (see Methods). Comparison shows that the variability in the SCEN experiment is considerably enhanced, and that the coupling explains a substantial fraction of the simulated future temperature variability (compare Fig. 1b and Fig. 1d). Overall, the effect of soil-moisture–temperature coupling in SCEN amounts to about two-thirds of the climate-change signal (Fig. 1e, f). Figure 1g and h displays a splitting of the total climate-change-induced variability change (Fig. 1e) into respective contributions owing to changes in external factors (atmospheric circulation, sea surface temperatures; Fig. 1g) and changes in land–atmosphere coupling (Fig. 1h; see Supplementary Discussion 2 for details). Disentangling the two signals shows that external effects are mostly restricted to France (Fig. 1g), while variability increases due to land–atmosphere coupling dominate in central and eastern Europe (Fig. 1h). Thus, changes in atmospheric circulation alone are unable to induce the projected increase in temperature variability, and the role of land–atmosphere coupling needs to be accounted for.

These results led us to ask why we find such a large impact of land–atmosphere coupling for future temperature variability in Europe, when the GLACE study (for present climate) did not¹¹. Could this be related to climate-change-induced modifications in land–atmosphere coupling characteristics? To address this question, we investigate the land–atmosphere coupling strength in CTL and SCEN using two different measures displayed in Fig. 2 (see Methods). The results obtained for these two measures are qualitatively similar and present two interesting features: (1) there is a shift of the region of highest soil-moisture–temperature coupling from the Mediterranean (CTL) to central and eastern Europe (SCEN); and (2) although

¹Institute for Atmospheric and Climate Science, ETH Zurich, Universitätsstrasse 16, 8092 Zurich, Switzerland.

the absolute impact of land–atmosphere coupling on temperature variability is substantially larger in SCEN (Fig. 1b versus Fig. 1a), the fraction of variance explained by land–atmosphere coupling in the Mediterranean, and in central and eastern Europe, respectively, increases only slightly from CTL to SCEN (about two-thirds of the total summer temperature variance in both cases).

We considered possible reasons for the differences in soil–moisture–temperature coupling in the Mediterranean region compared to the GLACE experiment. Such differences could be due to: (1) the higher resolution of our experiments (56 km compared to 1.8° – 3.8° or ~ 100 – 400 km); (2) representation of interannual variations in sea surface temperatures (SSTs; see Methods); (3) differences in model sensitivity, presumably due to parameter choices²¹.

Further analysis based on a large set of RCM and GCM experiments from the European project PRUDENCE²² and the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4)'s database, respectively, suggests that our model simulations are consistent with other RCM and GCM experiments (see Supplementary Discussion 1 and Supplementary Figures 1–3). To investigate this further, in Fig. 3 we compare the correlation between

evapotranspiration and temperature ($\rho_{ET, T_{2m}}$) in our simulations and in the investigated GCMs. This can be seen as a reverse measure of soil–moisture–temperature coupling, as negative correlations point to a strong control of soil moisture upon evapotranspiration and temperature (while positive correlations generally point to a strong atmospheric control on evapotranspiration; see Methods). The values of $\rho_{ET, T_{2m}}$ for our RCM experiments show (Fig. 3a, b) an excellent agreement with the measures of soil–moisture–temperature coupling strength (Fig. 2). Figure 3d–i displays $\rho_{ET, T_{2m}}$ for all GCMs (Fig. 3g–i), and for three GCMs (Fig. 3d–f) found to have high-quality circulation patterns in Europe²³. For the CTL time period, the large bands of negative $\rho_{ET, T_{2m}}$ in the GCMs show patterns clearly similar to those of the ‘hotspots’ of land–atmosphere coupling of the GLACE experiment^{7,11}. However, unlike the GLACE experiment but like our CTL experiment, they include the Mediterranean region (Fig. 2). Thus, it appears that the presence of interannual SST variations is the dominant mechanism explaining the difference between our CTL and the GLACE experiments. (This result is nevertheless qualitatively consistent with the GLACE experiment, because the Mediterranean region is a transitional zone between dry and wet climates).

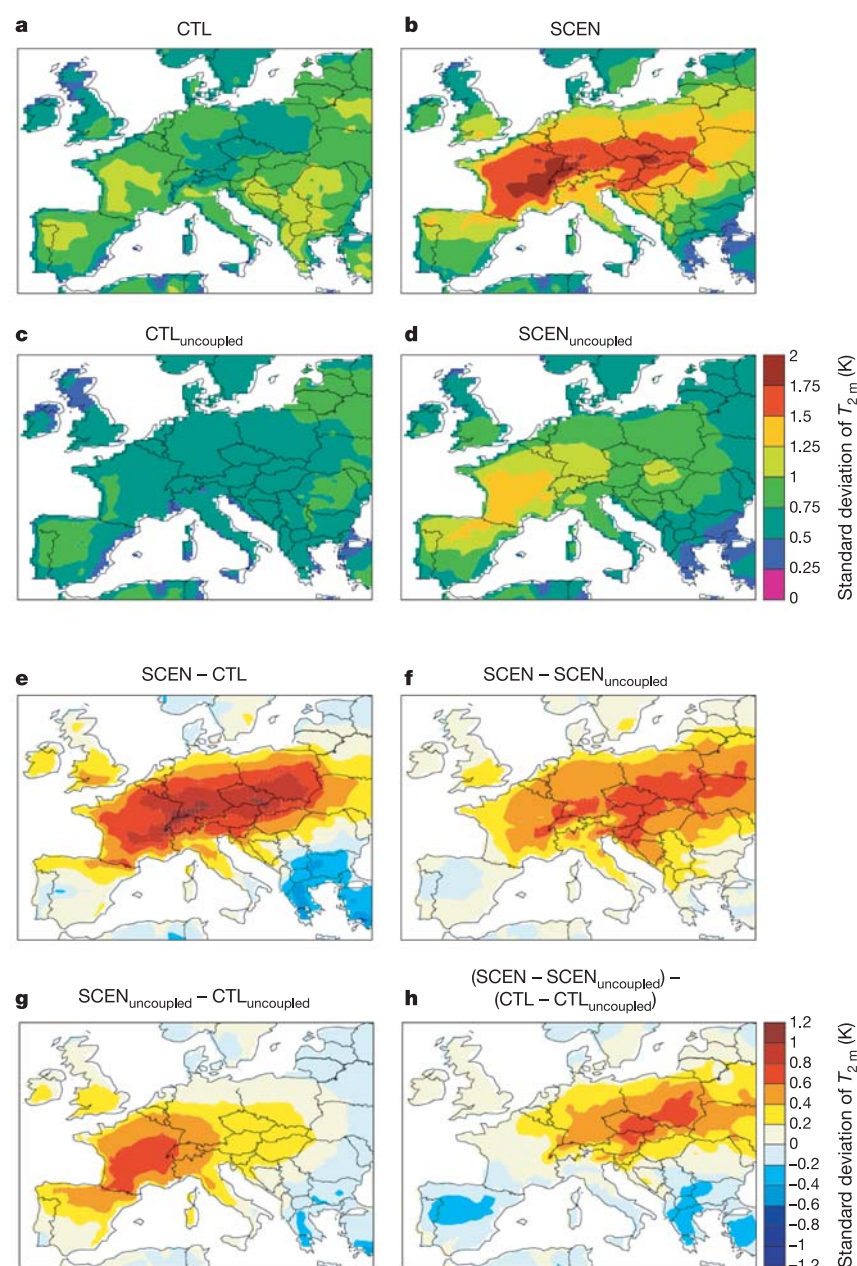


Figure 1 | Effects of land–atmosphere coupling on greenhouse-gas induced changes in interannual variability of summer two-metre temperature.

a–d, Standard deviation of summer (June–August) temperature in CTL (**a**), SCEN (**b**), CTL_{uncoupled} (**c**), and SCEN_{uncoupled} experiments (**d**). **e, f**, Differences between SCEN and CTL (**e**) and between SCEN and SCEN_{uncoupled} (**f**) experiments. **g, h**, Relative contributions to the climate-change signal of changes in external factors (circulation, sea surface temperatures), computed as SCEN_{uncoupled} – CTL_{uncoupled} (**g**), and changes in land–atmosphere coupling, computed as (SCEN – SCEN_{uncoupled}) – (CTL – CTL_{uncoupled}) (**h**). (See text and Supplementary Discussion 2 for more details).

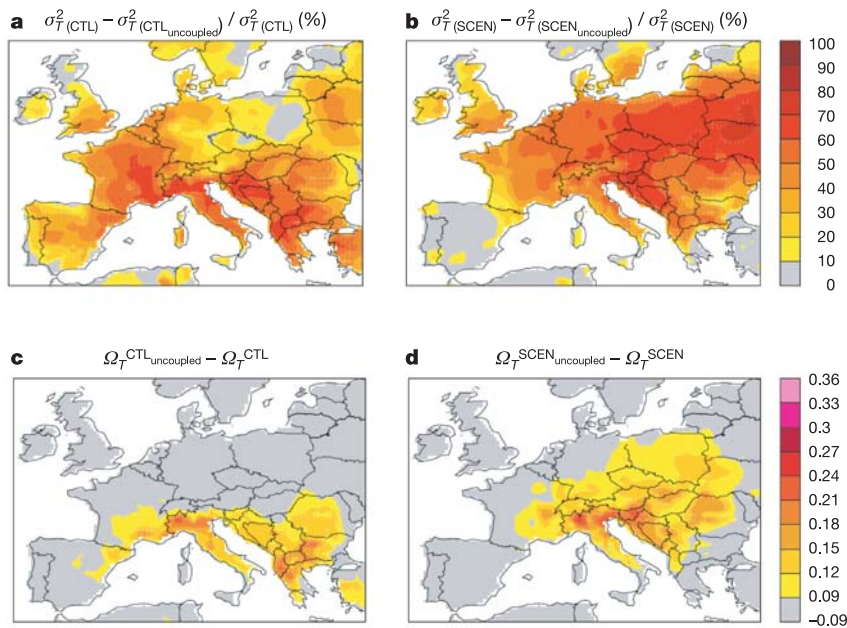


Figure 2 | Importance of soil-moisture-temperature coupling in present and future climate in terms of two different coupling diagnostics. a, b, Percentage of interannual summer temperature variance due to land-atmosphere coupling. **c, d,** Land-atmosphere coupling parameter for temperature computed in analogy to the GLACE experiment. (See Methods for a description of the coupling diagnostics).

Concerning the changes in $\rho_{ET, T2m}$ between present and future climate (Fig. 3c, f, i), a clear shift towards a more soil-moisture-controlled evapotranspiration regime in central Europe can be identified in the simulations. Figure 3 also shows that there are significant differences between GCMs regarding the magnitude of the effect (Fig. 3f, i), and that the RCM experiment (Fig. 3a, b) shows more geographical detail and on average a somewhat lower

atmospheric control on evapotranspiration than the GCM experiments (possibly owing to the models' resolution in the Alpine region). However, overall the analyses in Figs 2 and 3 are consistent and suggest that our findings are—at least qualitatively—not model specific.

Land-atmosphere coupling affects not only summer temperature variability, but also the whole water cycle in the simulations. Figure 4a–d presents a similar analysis to that of Fig. 1e–h, but for summer

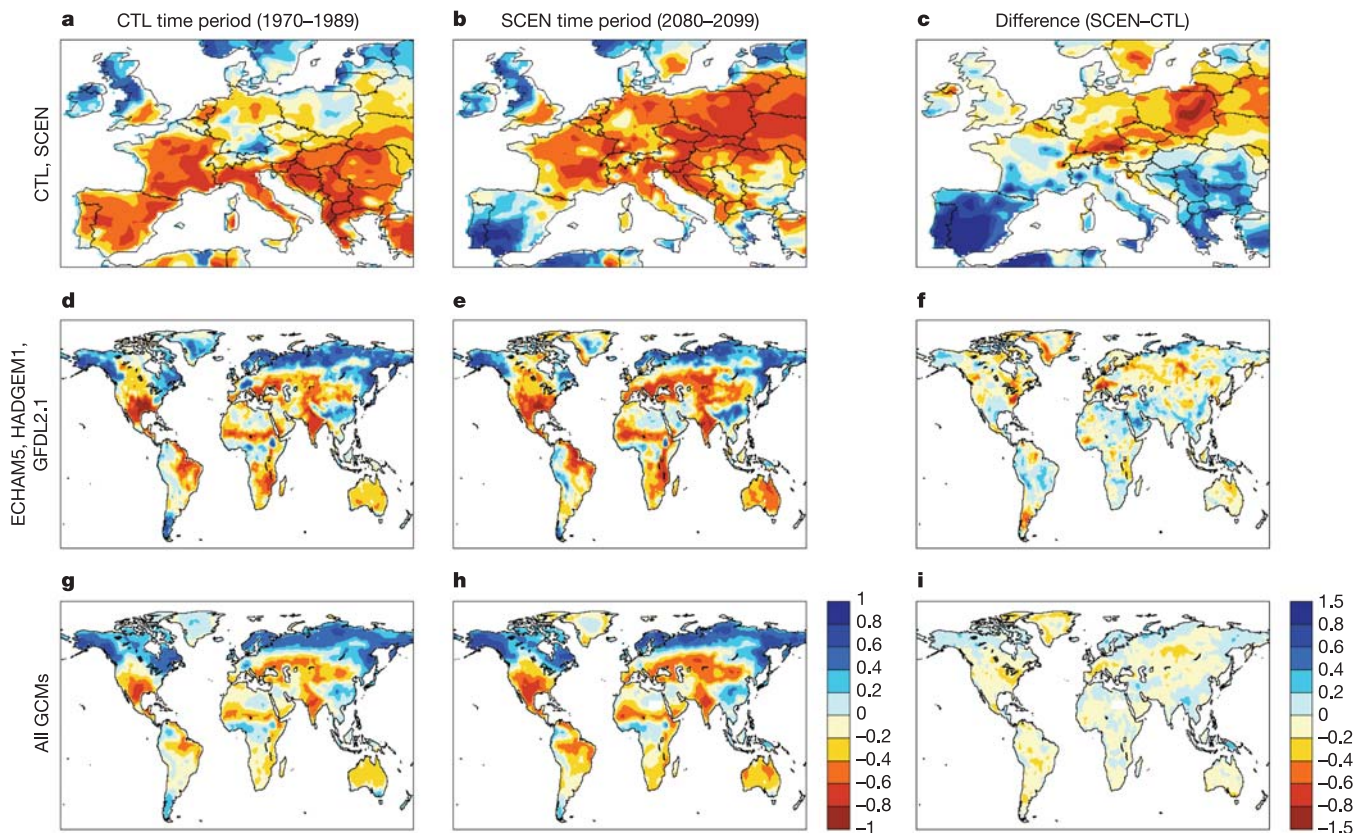


Figure 3 | Correlation of summer (June–August) evapotranspiration and temperature ($\rho_{ET, T2m}$) in the RCM and IPCC AR4 GCM experiments. The three columns display (from left to right) the CTL period (a, d, g), the SCEN period (b, e, h), and the change between the two periods (c, f, i).

The three rows show (from top to bottom): RCM results (a–c); mean of ECHAM5, HADGEM1 and GFDL2.1 GCMs (d–f); mean of all GCMs (g–i). (See Methods for a description of the analysed GCMs and the computation of $\rho_{ET, T2m}$).

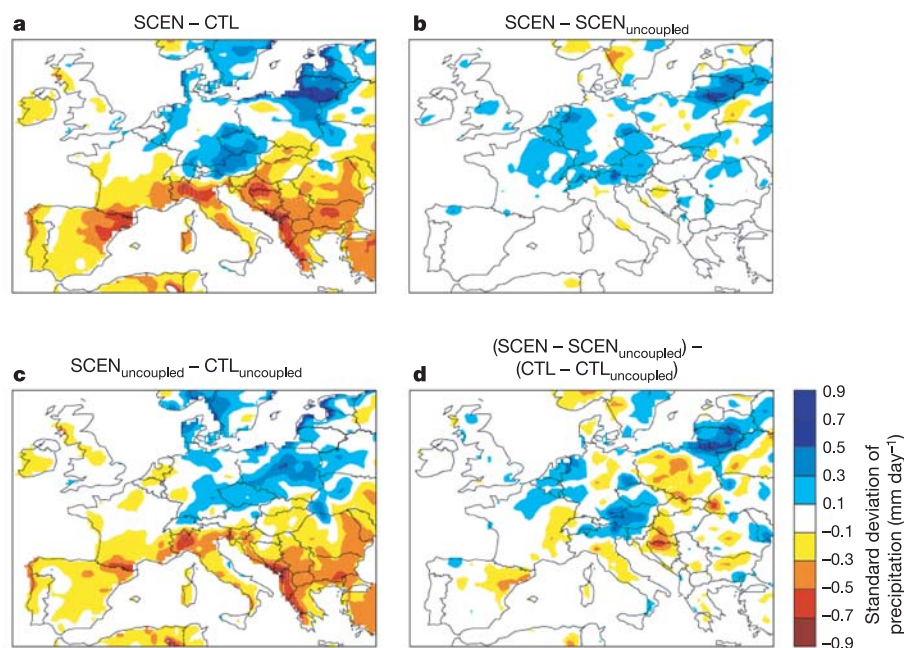


Figure 4 | Effects of land-atmosphere coupling on greenhouse-gas-induced changes in interannual variability of summer precipitation.

a, b, Differences in standard deviation of summer (June–August) precipitation: SCEN – CTL (**a**), SCEN – SCEN_{uncoupled} (**b**). **c, d**, Relative contributions to climate-change signal of changes in external factors

(circulation, sea surface temperatures), computed as SCEN_{uncoupled} – CTL_{uncoupled} (**c**), and changes in land–atmosphere coupling, computed as (SCEN – SCEN_{uncoupled}) – (CTL – CTL_{uncoupled}) (**d**). (See text and Supplementary Discussion 2 for more details).

precipitation variability. This comparison indicates that in most regions where an increase of precipitation variability is simulated (in particular in the Alpine region), the signal is at least partly caused by land–atmosphere coupling (see Supplementary Discussion 2 for more details). A computation similar to the one presented in Fig. 2b for temperature reveals that land–atmosphere coupling accounts again for about two-thirds of the total summer precipitation variance of the SCEN simulations in regions where increases of sizeable magnitude are simulated. Although there is generally more model uncertainty with regard to changes in precipitation than in temperature variability³, this result indicates another possibly key effect of land–atmosphere coupling.

In conclusion, the most striking result of our analysis is that land–atmosphere coupling is significantly affected by global warming and is itself a key player for climate change. Enhanced greenhouse gas concentrations lead to a northwards shift of climatic zones within the European continent. As a consequence, central and eastern Europe becomes a new transitional zone between dry and wet climates (similar to the Mediterranean region in the present climate), and thereby becomes susceptible to the effects of land–atmosphere coupling^{7,11}. These mechanisms can be identified both in our RCM experiment and in the analysed GCM simulations.

With regards to temperature variability (and associated heat-waves), our study does not contradict earlier reports that emphasized the role of atmospheric circulation changes^{5,15,18}. Rather, our results show that changes in both circulation and land-surface processes are necessary to explain the projected increase in variability. Here the impact of the circulation changes may be indirect, by imposing changes in the seasonal cycle of soil moisture^{24,25}, which in turn can lead to modified soil-moisture–temperature coupling characteristics. Finally, this investigation reveals how profoundly greenhouse gas forcing may affect the functioning of the regional climate system and the role of land-surface processes. At present, measurements of soil moisture²⁶ and evapotranspiration are extremely scarce (particularly in Europe) and do not allow an accurate assessment of these gradual changes. Therefore, a better monitoring of the terrestrial water and energy cycles in mid-latitude regions would be invaluable to further

our understanding of these processes and changes. Ultimately, it could also contribute to improved seasonal forecasting, and thereby help us to cope with the consequences of climate change.

METHODS

RCM climate-change experiments. The climate-change scenario is based on the SRES A2 transient greenhouse gas scenario as specified by the IPCC³¹. The scenario computations involve three numerical models: the low-resolution HadCM3 global coupled atmosphere–ocean GCM²⁷, the intermediate-resolution HadAM3H atmospheric GCM²⁸, and the CHRM limited-area high-resolution RCM²⁹. For details please refer to previous publications^{1,3}. A summary of the experimental set-up is given in Supplementary Table 1.

GCM climate-change experiments. The simulations are taken from the IPCC AR4 PCMDI data bank (http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php; 20C3M for present-climate and SRES A2 for future-climate experiments). A total of 12 GCM experiments with corresponding variables for the two periods are analysed: ECHAM5*, HADGEM1*, GFDL2.1*, CCCMA_T47, MIROC_MED, GISS, HADCM3, INM-CM, IPSL, MRI, NCAR_CCSM, NCAR_PCM. The models with asterisks (*) were recently identified as effectively representing circulation patterns in the northern mid- and high latitudes and in Europe²³ (the GCMs MIROC_HI and CCCMA_T63 are not available for the SRES A2 scenario period). Some of the analysis in Fig. 3 and the Supplementary Information refers to these three GCMs.

Computation of interannual summer variability of climate variables. The interannual summer variability of temperature and precipitation (Figs 1, 4)—and other climate variables presented in the Supplementary Information—is defined here as the standard deviation of the June–August mean values. Trend-induced inflation of the standard deviation (and variance, Fig. 2a, b) is removed using linear detrending³⁰. The detrending has some quantitative effect on the results but no significant qualitative impact.

Computation of soil-moisture–temperature coupling parameters. The coupling parameters we used are: the variance analysis, the GLACE-type coupling strength parameter, and the correlation between summer temperature and evapotranspiration.

Variance analysis. Figure 2a and b displays (for CTL and SCEN) the percentage of the interannual variance of June–August mean summer temperature that can be explained by land–atmosphere coupling, estimated as:

$$\frac{\sigma_{T(\text{coupled})}^2 - \sigma_{T(\text{uncoupled})}^2}{\sigma_{T(\text{coupled})}^2}$$

GLACE-type coupling strength parameter. The soil-moisture–temperature coupling strength parameter for present-day ($\Omega_T^{\text{CTLuncoupled}} - \Omega_T^{\text{CTL}}$) and future-climate ($\Omega_T^{\text{SCENuncoupled}} - \Omega_T^{\text{SCEN}}$) conditions displayed in Fig. 2c and d is computed in analogy to the approach followed in the GLACE experiment^{7,11}, with adaptations for a 20-year analysis sample. We consider here detrended timeseries of 6-day mean temperature values for June–August (neglecting the first 6-day interval to account for the same period as GLACE, that is, 14 values per summer) for 20 summers (corresponding to the 20 years of simulations considered). First the Ω_T^{CTL} , $\Omega_T^{\text{CTLuncoupled}}$, Ω_T^{SCEN} and $\Omega_T^{\text{SCENuncoupled}}$ values are computed as follows:

$$\Omega_T = \frac{20\sigma_T^2 - \sigma_T^2}{19\sigma_T^2}$$

where σ_T^2 is the standard deviation of 6-day mean temperature computed from all values available within the respective simulations (that is, 280 values in total) and σ_T^2 is the standard deviation of the 6-day mean temperature in the 20-year-average time series (that is, 14 values in total). The Ω_T values are estimates of the degree of interannual similarity in each experiment, so the values ($\Omega_T^{\text{CTLuncoupled}} - \Omega_T^{\text{CTL}}$) and ($\Omega_T^{\text{SCENuncoupled}} - \Omega_T^{\text{SCEN}}$) represent the extent to which the removal of interannual variability of soil moisture increases the interannual similarity (or decreases the interannual variability) of the simulations. For more details please refer to refs 7 and 11.

We note a few significant differences in the set-up of the present study compared to these earlier analyses^{7,11}. First, our ‘ensemble members’ are simulations for 20 individual summers corresponding to differing SST and atmospheric conditions, whereas the GLACE ensemble members are 16 simulations corresponding to one set of SST conditions (1994) but differing atmospheric conditions. Second, our ‘uncoupled’ simulations investigate the impact of prescribing soil moisture to the climatology, that is, two effects are included: (1) it implies the same soil moisture evolution for each year (or each ‘ensemble member’), which induces a zero interannual variation of soil moisture; (2) the use of the climatological soil-moisture evolution in each summer induces a damped intra-annual variation of soil moisture. In comparison, the GLACE experiment uses the soil-moisture evolution of one ensemble member in the ‘uncoupled’ simulations, which thus only investigates the impact of zero inter-member differences in soil-moisture evolution but does not damp the intra-annual evolution of soil moisture. The first point is relevant for the possible combination of SST-induced circulation changes with local amplification through land–atmosphere feedbacks; the second point may be more relevant for soil-moisture–precipitation rather than soil-moisture–temperature feedbacks, because increases in precipitation variability are expected to be related to increases in both intra-annual as well as interannual climate variability, while increases in temperature variability (in particular, linked to more frequent heatwaves) are expected to be more typically associated with increases in interannual climate variability.

Correlation between summer temperature and evapotranspiration. Figure 3a, b, d, e, g and h shows for the analysed RCM and GCM experiments the correlation between June–August mean summer temperature and evapotranspiration $\rho_{\text{ET}, T_{2\text{m}}}$ for the 20 years considered of the CTL and SCEN time periods (computed from detrended timeseries). We note that this measure of coupling is only meaningful in regions where evapotranspiration is not exceedingly small (for example, the Sahara in the GCM simulations or Spain in the RCM SCEN experiment).

Received 12 December 2005; accepted 20 July 2006.

- Schär, C. *et al.* The role of increasing temperature variability in European summer heatwaves. *Nature* **427**, 332–336 (2004).
- Giorgi, F., Bi, X. & Pal, J. S. Mean, interannual variability and trends in a regional climate change experiment over Europe. II. Climate change scenarios (1971–2100). *Clim. Dyn.* **23**, 839–858 (2004).
- Vidale, P. L., Lüthi, D., Wegmann, R. & Schär, C. European climate variability in a heterogeneous multi-model ensemble. *Clim. Change* (in the press).
- Räisänen, J. CO₂-induced changes in interannual temperature and precipitation variability in 19 CMIP2 experiments. *J. Clim.* **15**, 2395–2411 (2002).
- Meehl, G. A. & Tebaldi, C. More intense, more frequent, and longer lasting heat waves in the 21st century. *Science* **305**, 994–997 (2004).
- Lenderink, G., van Ulden, A., van den Hurk, B. & Van Meijgaard, E. Summertime interannual temperature variability in an ensemble of regional model simulations: analysis of surface energy budget. *Clim. Change* (in the press).
- Koster, R. D. *et al.* Regions of strong coupling between soil moisture and precipitation. *Science* **305**, 1138–1140 (2004).
- Betts, A. K. Understanding hydrometeorology using global models. *Bull. Am. Met. Soc.* **85**, 1673–1688 (2004).
- Schär, C., Lüthi, D., Beyerle, U. & Heise, E. The soil-precipitation feedback: A

- process study with a regional climate model. *J. Clim.* **12**, 722–741 (1999).
- Eltahir, E. A. B. A soil moisture-rainfall feedback mechanism. 1. Theory and observations. *Water Resour. Res.* **34**, 765–776 (1998).
- Koster, R. D. *et al.* GLACE: The Global Land–Atmosphere Coupling Experiment. Part 1. Overview. *J. Hydrometeor.* **7**, 590–610 (2006).
- Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M. & Wanner, H. European seasonal and annual temperature variability, trends, and extremes since 1500. *Science* **303**, 1499–1503 (2004).
- Stott, P. A., Stone, D. A. & Allen, M. R. Human contribution to the European heatwave of 2003. *Nature* **432**, 610–614 (2004).
- Christensen, J. H. & Christensen, O. B. Severe summertime flooding in Europe. *Nature* **421**, 805–806 (2003).
- Pal, J. S., Giorgi, F. & Bi, X. Consistency of recent European summer precipitation trends and extremes with future regional climate projections. *Geophys. Res. Lett.* **31**, L13202 (2004).
- Milly, P. C. D., Dunne, K. A. & Vecchia, A. V. Global pattern of trends in streamflow and water availability in a changing climate. *Nature* **438**, 347–350 (2005).
- Frei, C., Schöll, R., Fukutome, S., Schmidli, J. & Vidale, P. L. Future change of precipitation extremes in Europe: An intercomparison of scenarios from regional climate models. *J. Geophys. Res.* **111**, D06105 (2006).
- Ogi, M., Yamazaki, K. & Tachibana, Y. The summer northern annular mode and abnormal summer weather in 2003. *Geophys. Res. Lett.* **32**, L04706 (2005).
- Sutton, R. T. & Hodson, D. L. R. Atlantic ocean forcing of North American and European summer climate. *Science* **309**, 115–118 (2005).
- Koster, R. D., Suarez, M. J. & Heiser, M. Variance and predictability of precipitation at seasonal-to-interannual time scales. *J. Hydrometeor.* **1**, 26–46 (2000).
- Lawrence, D. M. & Slingo, J. M. Weak land–atmosphere coupling in HadAM3: role of soil moisture variability. *J. Hydrometeor.* **6**, 670–680 (2005).
- Christensen, J. H., Carter, T. R. & Giorgi, F. PRUDENCE employs new methods to assess European climate change. *Eos* **83**, 147 (2002).
- van Ulden, A. O. & van Oldenborgh, G. J. Large-scale atmospheric circulation biases and changes in global climate simulations and their importance for climate change in Central Europe. *Atmos. Chem. Phys.* **6**, 863–881 (2006).
- Findell, K. L. & Delworth, T. L. A modeling study of dynamic and thermodynamic mechanisms for summer drying in response to global warming. *Geophys. Res. Lett.* **32**, L16702 (2005).
- Seneviratne, S. I., Pal, J. S., Eltahir, E. A. B. & Schär, C. Summer dryness in a warmer climate: a process study with a regional climate model. *Clim. Dyn.* **20**, 69–85 (2002).
- Robock, A. *et al.* The global soil moisture data bank. *Bull. Am. Meteorol. Soc.* **81**, 1281–1299 (2000).
- Johns, T. C. *et al.* Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios. *Clim. Dyn.* **20**, 583–612 (2003).
- Pope, D. V., Gallani, M., Rowntree, R. & Stratton, A. The impact of new physical parameterizations in the Hadley Centre climate model HadAM3. *Clim. Dyn.* **16**, 123–146 (2000).
- Vidale, P. L., Lüthi, D., Frei, C., Seneviratne, S. I. & Schär, C. Predictability and uncertainty in a regional climate model. *J. Geophys. Res.* **108** (D18), 4586, (2003).
- Scherrer, S. C., Appenzeller, C., Liniger, M. A. & Schär, C. European temperature distribution changes in observations and climate change scenarios. *Geophys. Res. Lett.* **32**, L19705 (2005).
- Nakicenovic, N. *et al.* *IPCC Special Report on Emissions Scenarios* (eds Nakicenovic, N. & Swart, R.) (Cambridge Univ. Press, Cambridge, UK, 2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The computations were performed on the computing facilities of ETH Zurich and the Swiss Center for Scientific Computing (CSCS). We thank the Hadley Centre, UK, for providing access to their climate-change simulations, the PRUDENCE team for earlier interactions and model inter-comparisons (unperturbed simulations), the international modelling groups for providing the IPCC AR4 data through PCMDI, and the GFDL modelling group for access to the GFDL evapotranspiration fields. We acknowledge the JSC/CLIVAR Working Group on Coupled Modelling (WGCM) and their Coupled Model Intercomparison Project (CMIP) and Climate Simulation Panel for organizing the IPCC model data analysis activity, and the IPCC WG1 TSU for technical support. We also thank our colleagues for comments on the manuscript. This research was supported by the Swiss National Science Foundation (NCCR Climate) and by the sixth Framework Programme of the European Union (project ENSEMBLES).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.I.S. (sonia.seneviratne@env.ethz.ch).

LETTERS

The tumour suppressor Hippo acts with the NDR kinases in dendritic tiling and maintenance

Kazuo Emoto^{1†}, Jay Z. Parrish¹, Lily Yeh Jan¹ & Yuh-Nung Jan¹

Precise patterning of dendritic fields is essential for neuronal circuit formation and function, but how neurons establish and maintain their dendritic fields during development is poorly understood. In *Drosophila* class IV dendritic arborization neurons, dendritic tiling, which allows for the complete but non-overlapping coverage of the dendritic fields^{1–3}, is established through a 'like-repels-like' behaviour of dendrites mediated by Tricornered (Trc), one of two NDR (nuclear Dbp2-related) family kinases in *Drosophila*^{4–7}. Here we report that the other NDR family kinase, the tumour suppressor Warts/Lats (Wts)^{8–10}, regulates the maintenance of dendrites; in *wts* mutants, dendrites initially tile the body wall normally, but progressively lose branches at later larval stages, whereas the axon shows no obvious defects. We further provide biochemical and genetic evidence for the tumour suppressor kinase Hippo (Hpo)^{11–15} as an upstream regulator of Wts and Trc for dendrite maintenance and tiling, respectively, thereby revealing important functions of tumour suppressor genes of the Hpo signalling pathway in dendrite morphogenesis.

Dendritic arborization patterns are critical to a neuron's ability to receive and process impinging signals. Whereas neurons normally maintain the gross morphology of their dendrites^{16,17}, cortical neurons of Down's syndrome patients gradually lose dendritic branches after initially forming normal dendritic fields^{18,19}. Thus, neurons appear to have separate mechanisms for establishment and maintenance of their dendritic fields.

Dendritic tiling is an evolutionarily conserved mechanism for neurons of the same type to ensure complete but non-redundant coverage of dendritic fields^{1–3}. In the mammalian visual system, for instance, dendrites of each retinal ganglion cell type cover the entire retina with little overlap, like tiles on a floor¹. In *Drosophila*, the dendritic arborization sensory neurons can be divided into four classes (I–IV) based on their dendrite morphology⁴, and the dendritic field of class IV dendritic arborization neurons is shaped, in part, through a like-repels-like tiling behaviour of dendrite terminals^{5,6}. We recently identified the NDR family kinase Trc and its activator Furry (Fry) as essential regulators of dendritic tiling and branching of class IV dendritic arborization neurons⁷. These proteins are evolutionarily conserved and probably serve similar functions in neurons of different organisms^{7,10,20,21}.

In addition to Trc, *Drosophila* has one other NDR family kinase, Wts, which is a tumour suppressor protein that functions in the coordination of cell proliferation and cell death in flies^{8,9} and vertebrates²². To uncover the cell-autonomous functions of Wts in neurons, we used MARCM (mosaic analysis with a repressive cell marker) to generate mCD8–GFP-labelled *wts* clones in a heterozygous background²³. Wild-type class IV neurons elaborate highly branched dendrites that cover essentially the entire body wall. Compared to wild-type ddaC (dorsal dendrite arborization neuron C) neurons (Fig. 1a),

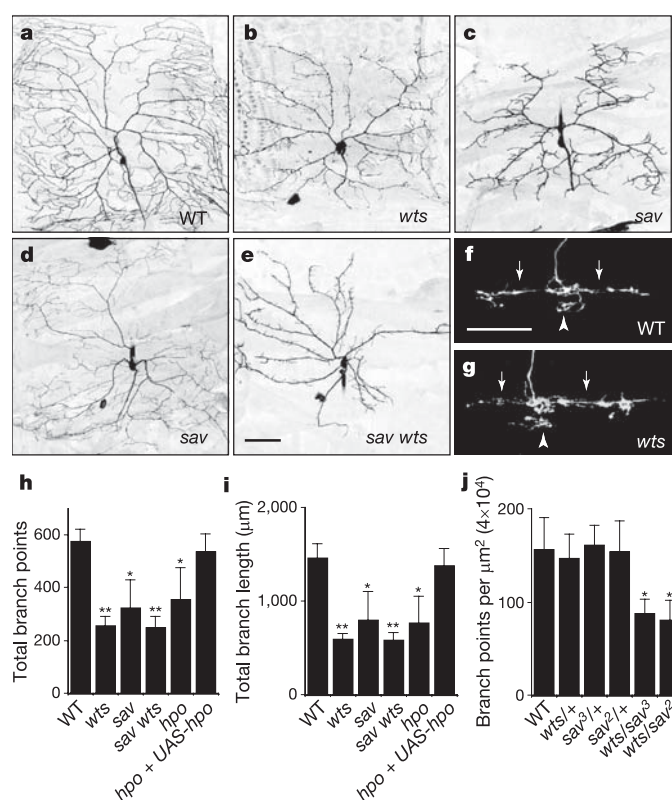


Figure 1 | *wts* and *sav* mutants have similar defects in dendrite morphogenesis. **a–e**, Compared to wild-type (WT) clones (**a**), mutant clones of ddaC class IV neurons lacking the function of *wts* (**b**) or *sav* (**c, d**), or both (**e**), are defective in dendritic coverage. Anterior is left and dorsal is up. Genotypes: *FRT82B, wts^{lastX1}* (**b**); *FRT82B, sav³* (**c, d**); *FRT82B, sav³, wts^{lastX1}* (**e**). **f, g**, ddaC axon terminals with bifurcated ipsilateral branches (arrows) and a short commissural branch (arrow heads) in wild-type (**f**) and *wts* mutant clones (**g**). **h, i**, Quantification of the total branch points (**h**) and total dendrite length (**i**) of MARCM clones (wild type, *n* = 7; *wts*, *n* = 10; *sav*, *n* = 12; *sav wts*, *n* = 4; *hpo*, *n* = 10; *hpo + UAS-hpo*, *n* = 4). Genotypes: *wts^{lastX1} (wts)*, *sav³ (sav)*, *sav³, wts^{lastX1} (sav wts)*, *hpo^{mgh4} (hpo)*, and *hpo^{mgh4}; ppk-Gal4/UAS-hpo (hpo + UAS-hpo)*. Error bars show mean ± s.d. Double asterisk, *P* < 0.001; single asterisk, *P* < 0.005 (Student's *t*-test). **j**, Trans-heterozygous combinations of *wts* and *sav* mutants phenocopy the dendrite defects of homozygous *wts* or *sav* mutants. Genotypes: *wts^{lastX1}/+* (*wts/+*), *wts^{lastX1}/sav³* (*wts/sav³*), *wts^{lastX1}/sav²* (*wts/sav²*). Error bars show mean ± s.d., *n* = 25, *P* < 0.01 (Student's *t*-test). The scale bars in **e** and **f** represent 50 μm (**a–e**) and 20 μm (**f, g**), respectively.

¹Howard Hughes Medical Institute, Departments of Physiology, Biochemistry, and Biophysics, University of California San Francisco, San Francisco, California 94143-0725, USA.

[†]Present address: Neural Morphology Laboratory, National Institute of Genetics, Yata 1111, Mishima, Shizuoka 411-8540, Japan.

wts clones showed a severe and highly penetrant simplification of dendritic trees (Fig. 1b), with significantly reduced number (wild type, 575.1 ± 45.3 , $n = 7$; *wts*, 255.6 ± 35.8 , $n = 10$; $P < 0.001$) and length (wild type, $1,457.0 \pm 156.9 \mu\text{m}$, $n = 7$; *wts*, $590.4 \pm 65.0 \mu\text{m}$, $n = 10$; $P < 0.001$) of dendritic branches, and hence a greatly reduced dendritic field (Fig. 1b, h, i).

In contrast to the severe dendritic defects caused by loss of Wts function, *wts* mutant ddaC axons entered the ventral nerve cord at the appropriate position and showed arborization patterns very similar to wild-type controls (Fig. 1f, g and Supplementary Fig. 1a–f), with their axons terminating on the innermost fascicle and sending ipsilateral branches anteriorly and posteriorly (Fig. 1f, g; arrows) and sometimes also a collateral branch towards the midline (Fig. 1f, g; arrowheads). Thus, Wts seems to have a crucial role in dendrite-specific morphogenesis in post-mitotic neurons.

In proliferating cells, Wts is part of a signalling complex for tumour suppression that includes the adaptor protein Salvador (Sav) and the serine/threonine kinase Hpo. We next examined *sav* mutant ddaC MARCM clones and observed dendritic defects similar to *wts* MARCM clones (Fig. 1c, d, h, i). In severely affected clones (3 of 15 clones), most of the high-order branches were missing (Fig. 1c), whereas moderately affected clones (12 of 15 clones) exhibited a partial loss of their fine branches and major branches (Fig. 1d).

To confirm that Wts and Sav function in the same pathway, we tested for a genetic interaction between *wts* and *sav* in regulating dendrite morphogenesis. Whereas heterozygous *wts* or *sav* mutants had no obvious dendritic phenotype, trans-heterozygous combinations of *wts* and *sav* alleles resulted in simplified dendrites similar to moderately affected *sav* clones (Fig. 1j). Furthermore, *sav wts* double mutant clones showed a severe dendrite defect comparable to *wts* mutant clones (Fig. 1e). Thus, Wts and Sav most probably function together in class IV neurons to regulate dendrite morphogenesis.

The dendritic phenotypes of *wts* mutants and *sav* mutants might result from defects in branch formation and/or elongation, or loss of normally formed dendrites. We therefore examined ddaC dendrites at different time points of larval development using the *pickpocket-EGFP* reporter, which is specifically expressed in class IV dendritic arborization neurons⁵. Wild-type ddaC neurons elaborated primary and secondary dendritic branches by 24–28 h after egg laying, but large regions of the body wall were not yet covered by dendrites (Fig. 2a). By 48–52 h after egg laying, the major branches reached the dorsal midline, and the open spaces between major branches were filled with fine branches, resulting in complete dendritic coverage of the body wall (Fig. 2b). This tiling of dendrites persisted throughout the rest of larval development (Fig. 2c). In *wts* and *sav* mutants, ddaC dendrites were indistinguishable from those of wild-type controls at 24–28 h after egg laying (Fig. 2d, g, j). By 48–52 h after egg laying, *wts* and *sav* dendrites tiled the body wall as in wild type (Fig. 2e, h, j). During the next 24 h, however, dendrites of *wts* and *sav* mutants no longer tiled the body wall (Fig. 2f, i, j). Therefore, *wts* and *sav* seem to be required for maintenance of the already established tiling of dendrites.

The loss of dendrites was further documented in live mutant larvae imaged for 30 h starting in early second instar larvae (48–50 h after egg laying). In wild-type larvae, ddaC dendrites grew steadily; the number of terminal branches increased by $23.0 \pm 3.2\%$ ($n = 5$) over this time period (Supplementary Fig. 2a–c). By contrast, dendrites of *wts* and *sav* mutants gradually lost their fine branches (decrease of $27.5 \pm 5.3\%$ ($n = 4$) and $31.5 \pm 6.3\%$ ($n = 7$), respectively) as well as some of the major branches by 78–80 h after egg laying (Supplementary Fig. 2d–i).

Class-IV-neuron-specific expression of *wts* and *sav* largely rescued the dendritic phenotype of *wts* and *sav* mutants, respectively (Fig. 2j), confirming that Wts and Sav act cell autonomously in class IV neurons. Furthermore, no detectable defect in patterning of the epidermis (anti-Armadillo antibody) or muscle (Tropomyosin:GFP reporter) was observed in *wts* or *sav* mutant third instar larvae (data not shown). Taken together, these results indicate that the Wts/Sav

signalling pathway functions in class IV neurons to maintain dendritic arborizations.

Wts kinase activity is regulated, at least in part, by the Ste20-like serine/threonine kinase Hpo (refs 11–15). Indeed, *ddaC* clones mutant for *hpo* exhibited simplified dendritic trees in third instar larvae, similar to *wts* or *sav* mutant clones (Fig. 1h, i, 3a, b), but showed more extensive dendritic arborizations in earlier larval stages (second to early third instar) (Fig. 3c), consistent with the involvement of Hpo in the maintenance of dendrites. Notably, in *hpo* mutant clones at earlier developmental stages, dendritic branches were often found to overlap (Fig. 3c, g). Both the dendritic tiling and maintenance phenotypes were rescued by *hpo* expression in MARCM clones (Fig. 1h, i, 3g), consistent with the cell-autonomous function of Hpo in class IV neurons. Because this tiling defect in *hpo* mutant clones was similar to the tiling defects of *trc* mutant clones⁷, we next investigated whether *hpo* could genetically interact with *trc* to regulate dendritic tiling. Compared with wild-type controls (Fig. 3d and Supplementary Fig. 3a, b), trans-heterozygous combinations of *trc* and *hpo* exhibited obvious iso-neuronal (Fig. 3f, arrows, 3h) as well as hetero-neuronal tiling defects (Supplementary Fig. 3c, d), whereas *wts* and *hpo* trans-heterozygotes displayed simplified dendrites similar to *wts* mutants (Fig. 3e and Supplementary Fig. 4).

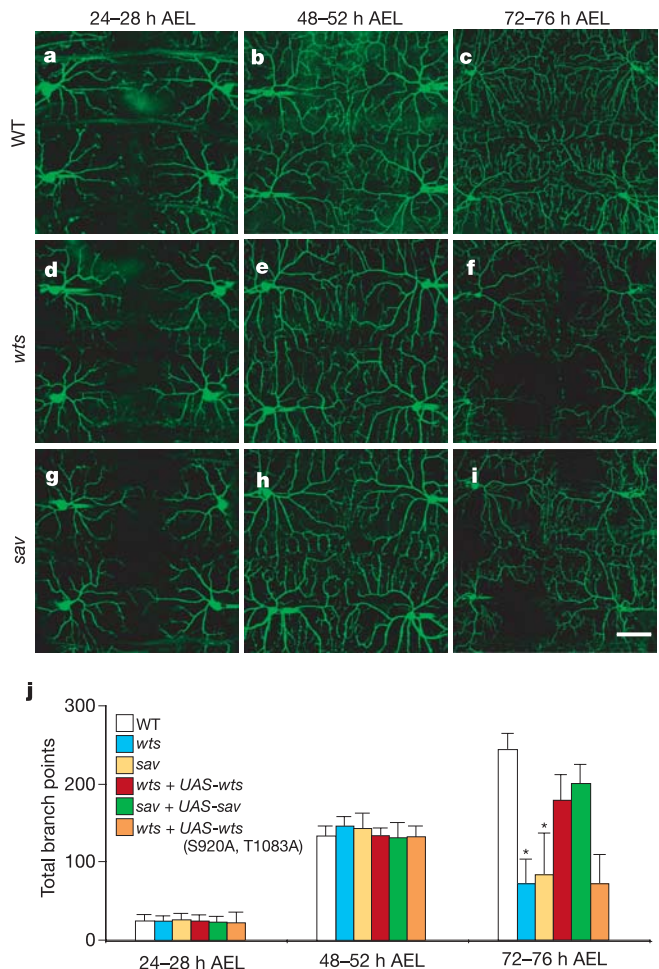


Figure 2 | Late-onset dendritic loss in *wts* and *sav* mutants in late larval stages. **a–i**, Live images of wild-type (**a–c**), *wts* mutant (**d–f**) and *sav* mutant (**g–i**) dendrites at different times after egg laying (AEL). The scale bar in **i** represents 10 μm (**a**, **d**, **g**), 15 μm (**b**, **e**, **h**) and 25 μm (**c**, **f**, **i**), respectively. **j**, Quantification of total branch points over 48 h. Error bars show mean $\pm$ s.d., $n = 15$. Asterisk, $P < 0.01$ (Student's *t*-test). Genotypes for rescue: *ppk-Gal4/UAS-wts*; *wts*^{P1}, *ppk-EGFP* (*wts* + *UAS-wts*), *ppk-Gal4/UAS-sav*; *sav*³, *ppk-EGFP* (*sav* + *UAS-sav*), and *ppk-Gal4/UAS-wts* (S920A, T1083A); *wts*^{P1}, *ppk-EGFP* (*wts* + *UAS-wts* (S920A, T1083A)).

These dendritic defects were consistently observed in multiple allelic combinations between *hpo* and *trc* or *wts* (Fig. 3h and Supplementary Fig. 4). In contrast, trans-heterozygous combinations of *trc* and *wts* showed no significant dendritic phenotypes (Fig. 3h and Supplementary Fig. 4). Furthermore, overexpression of wild-type Trc, but not Wts, in *hpo* MARCM clones partially rescued the dendritic tiling defects in class IV neurons (Fig. 3g). Thus, Hpo acts through Trc and Wts to regulate dendritic tiling and maintenance, respectively.

Not only did Hpo interact genetically with Trc and Wts, its physical association with these NDR kinases could be detected *in vivo*. When Flag-tagged Trc was expressed using a nervous-system-specific Gal4 driver, anti-Flag antibodies immunoprecipitated Trc together with Hpo (Fig. 4a). Similarly, Myc-tagged Wts co-immunoprecipitated with Hpo expressed in embryonic nervous systems (Fig. 4a). Hpo co-

immunoprecipitation appeared to be specific, because Misshapen, another Ste20-like kinase protein present in neurons²⁴, was not co-immunoprecipitated by anti-Flag or anti-Myc antibodies in similar experiments (data not shown). These results suggest that Hpo associates with Trc and Wts in the *Drosophila* nervous system.

To examine further the physical interaction between Trc and Hpo, we carried out analogous experiments in *Drosophila* S2 cells co-transfected with a haemagglutinin (HA)-tagged Trc construct and a Flag-tagged Hpo construct containing the full open reading frame, an amino-terminal fragment containing the kinase domain, or a carboxy-terminal fragment containing the regulatory domain¹⁵. As shown in Fig. 4b, full-length Hpo and the C-terminal portion of Hpo, but not the N-terminal fragment, were co-immunoprecipitated with Trc, suggesting that the C-terminal domain of Hpo is sufficient for Trc–Hpo complex formation.

Hpo physically interacts with Wts (refs 15, 25) and promotes Wts phosphorylation at multiple serine/threonine sites, including two sites^{11–15}, S920 and T1083 of *Drosophila* Wts, that appear to be necessary for Wts kinase activation²⁵. Indeed, Wts protein with mutations in the S920 and T1083 residues was unable to rescue the *wts* mutant dendritic phenotypes (Fig. 2j). Given that the corresponding phosphorylation sites in Trc (Fig. 4c) are critical for Trc activation as well as control of dendritic tiling and branching⁷, we wondered whether Hpo may promote Trc phosphorylation at the critical serine and/or threonine residue. Wild-type Hpo, but not catalytically inactive Hpo or the Misshapen kinase, led to substantial incorporation of ³²P-labelled phosphate into recombinant Trc (Fig. 4d and Supplementary Fig. 5) or Trc with a mutation at the S292 site (S292A), but not the T449A mutant form of Trc (Fig. 4d).

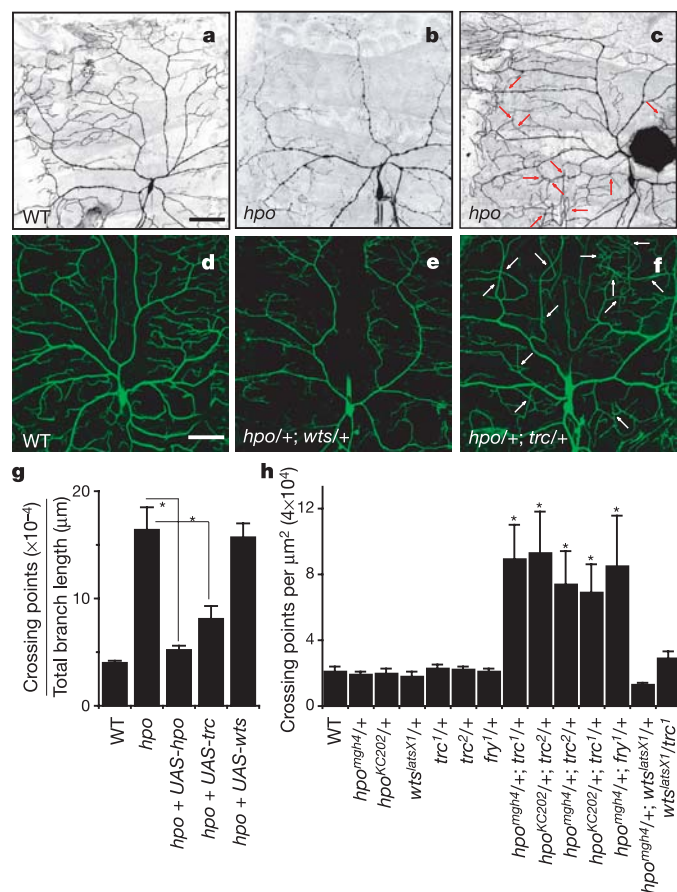


Figure 3 | *hpo* mutants display both dendritic tiling and maintenance defects. **a–c**, Dendrites of ddaC clones in late third instar larvae for wild-type control (**a**) and *hpo* mutants with phenotypes similar to *wts* and *sav* mutants (**b**), in contrast to *hpo* mutants at the late second instar larval stage with abnormal crossing of dendrites (red arrows) similar to the tiling phenotype of *trc* mutants (**c**). Genotype: *FRT42D*, *hpo*^{mgh4}. The scale bar in **a** represents 50 μm (**a–c**). **d–f**, Live images of ddaC dendrites in third instar larvae visualized by *ppk-EGFP* reporter in wild type (**d**), and trans-heterozygotes for *wts*^{lats} and *hpo*^{mgh4} (**e**), or for *trc*¹ and *hpo*^{mgh4} (**f**). Arrows in **f** indicate crossing points of dendritic branches. The scale bar in **d** represents 50 μm (**d–f**). **g**, Genetic rescue of dendritic tiling phenotypes of *hpo* MARCM clones in second instar larvae. (WT, *n* = 7; *hpo*, *n* = 10; *hpo* + UAS-*hpo*, *n* = 4; *hpo* + UAS-*trc*, *n* = 5; *hpo* + UAS-*wts*, *n* = 5). Error bars show mean ± s.d., *P* < 0.001 (Student's *t*-test). Genotypes: *w*, *elav-Gal4*, UAS-*mCD8-GFP*, *hs-flp*; *FRT42D*, *hpo*^{mgh4}/*FRT42D*, *tub-Gal80*, UAS-*hpo*+/+ (*hpo* + UAS-*hpo*); *w*, *elav-Gal4*, UAS-*mCD8-GFP*, *hs-flp*; *FRT42D*, *hpo*^{mgh4}/*FRT42D*, *tub-Gal80*, UAS-*trc*+/+ (*hpo* + UAS-*trc*); *w*, *elav-Gal4*, UAS-*mCD8-GFP*, *hs-flp*; *FRT42D*, *hpo*^{mgh4}/*FRT42D*, *tub-Gal80*, UAS-*wts*+/+ (*hpo* + UAS-*wts*). **h**, Quantification of crossing points per μm² (4×10^4) of dendritic branches at the second instar larval stage in ddaC neurons. Error bars show mean ± s.d., *n* = 15, *P* < 0.001 (Student's *t*-test).

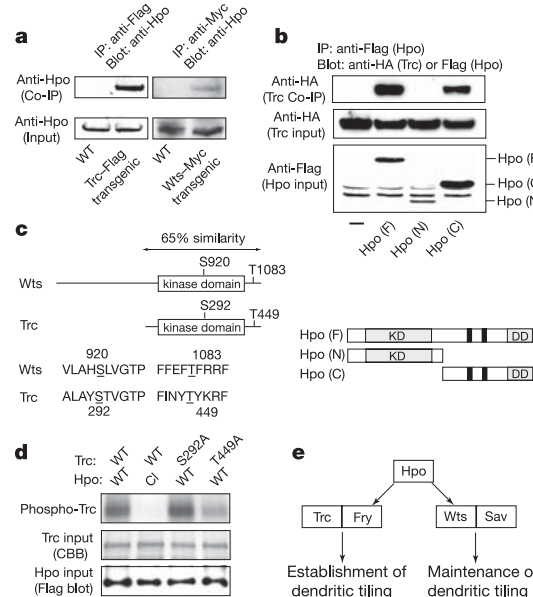


Figure 4 | Hpo binds to and phosphorylates Trc at a functionally critical threonine residue. **a**, Hpo was co-immunoprecipitated with neuronally expressed Trc–Flag (left panels) or Wts–Myc (right panels) from transgenic fly embryos as revealed by western blot analysis using anti-Hpo antibody. **b**, Immunoprecipitation via anti-Flag beads from S2 cell lysates revealing co-immunoprecipitation of Trc–HA with Flag-tagged full-length Hpo (F), or the C-terminal domain of Hpo (C), but not the N-terminal domain of Hpo (N). Schematic drawings are shown below. DD, dimerization domain; KD, kinase domain; black box, nuclear-export sequence. **c**, Diagram of Wts and Trc protein sequences and the alignments of their putative phosphorylation sites. **d**, *In vitro* kinase assays reveal that Hpo can phosphorylate Trc at T449, which is required for Trc regulation of dendritic tiling. Top panel shows ³²P incorporation. Middle and bottom panels show Trc and Hpo in each kinase reaction. CBB, Coomassie brilliant blue stain; CI, catalytically inactive Hpo kinase. **e**, Summary of the role of Hpo, Wts and Trc kinase signalling pathways in dendritic field establishment and maintenance.

Analogous results were obtained with Wts (Supplementary Fig. 6). These results support a model in which Hpo associates with and phosphorylates Trc and Wts at a critical threonine residue to regulate dendritic tiling and maintenance, respectively.

Both genetic and biochemical evidence reveals that Hpo regulates complementary aspects of dendrite development through two distinct downstream signalling pathways: the Trc kinase pathway for tiling and the Wts kinase pathway for maintenance (Fig. 4e). Our studies of class IV dendritic arborization neurons, together with the recent report that Wts signalling is required for cell fate specification of photoreceptor cells in *Drosophila* retina²⁶, demonstrate that the Wts signalling pathway is important for post-mitotic neurons. In proliferating cells, Wts phosphorylates Yorkie (Yki), a transcriptional co-activator, to regulate cell cycle and apoptosis in growing cells²⁷. However, Yki is dispensable for Hpo/Wts-mediated dendrite maintenance (Supplementary Fig. 7). Hpo probably functions as an upstream kinase for Trc, as well as Wts, in neurons by phosphorylating a functionally essential threonine (Fig. 4e), which may also be regulated by MST3, a Ste20-like kinase closely related to Hpo (ref. 28). Given the evolutionary conservation of known components in the Trc and Wts signalling pathways^{7,10,20,21}, it will be important to identify their relevant downstream targets and explore mechanisms that coordinate the establishment and maintenance of dendritic fields, and to determine the role of Trc and Wts signalling in the mammalian nervous system.

METHODS

Fly stocks. The following lines were used for the study: *wts^{latsX1}* (null)⁹, *wts^{P1}* (hypomorph)²⁶, *sav³* (null)¹¹, *sav²* (hypomorph)¹¹, *sav³ wts^{latsX1}* (a gift from N. Tapon), *hpo^{mgh4}* (null; from I. Hariharan), *hpo^{KC202}* (null)¹⁴, *trc¹* (null)⁷, *trc²* (null)⁷, *yki^{B5}* (null)²⁷, *yki^{B5} hpo⁴²⁻⁴⁷* (ref. 27), *UAS-trc-Flag* (ref. 7), and *UAS-wts-Myc* (ref. 29). For visualizing class IV dendrites, we used *pickpocket* (*ppk*)-EGFP (ref. 5), *wts^{P1}*, *ppk-EGFP* and *sav³*, *ppk-EGFP*. For genetic rescue experiments, we used *w/yw*; *ppk-Gal4/UAS-wts*; *wts^{P1}*, *ppk-EGFP* and *w*; *ppk-Gal4/UAS-sav*; *sav³*, *ppk-EGFP*.

MARCM analysis. MARCM analyses were performed as described previously with some modifications⁷. Briefly, to generate mosaic clones, *yw*; *wts^{latsX1}*, *FRT82B/TM6B*, *yw*; *sav³*, *FRT82B/TM3*, *yw*; *hpo^{mgh4}*, *FRT42D/CyO*, *w*; *p[FRT82B]* or *w*; *p[FRT42D]* flies were mated with *w*; *elav-Gal4*, *hsFLP*; *FRT82B*, *tub-Gal80/TM6B* or *w*; *elav-Gal4*, *hsFLP*; *FRT42D*, *tub-Gal80/CyO* flies. Embryos were collected for 2 h and aged for 3–5 h at 25 °C before being subjected to the following heat shock regime: 38 °C for 45 min, room temperature recovery for 30 min, and finally 38 °C for 45 min. The eggs were kept at 25 °C and larvae were examined for mutant clones and then dissected, fixed and stained with anti-mCD8 antibody (1:200 dilution; Caltag).

Kinase assay. Wild-type or mutant *trc* or *wts* cDNAs were subcloned into pGEX 5X-1 (Amersham) at *EcoRI* and *SalI* sites. Recombinant bacterial GST–Trc and GST–Wts proteins (kinase domain: 601–1105 amino acids) were purified using glutathione-sepharose (Amersham) and then cleaved by Factor Xa (NEB) to remove GST. Wts (kinase domain) and Trc recombinant proteins had auto-phosphorylation activity *in vitro*, therefore we used a catalytically inactivate Trc or Wts mutant as a substrate for the *in vitro* assay for Hpo kinase. Wild-type or catalytically inactive mutant of Hpo were expressed individually in *Drosophila* S2 cells. After immunoprecipitation of the proteins via their Flag tags, Hpo proteins were eluted from the beads by Flag peptides (Sigma). The *in vitro* kinase assays were carried out in kinase buffer (20 mM Tris-HCl, pH 7.4, 10 mM MgCl₂, 1 mM dithiothreitol, 100 μM ATP and 10 μCi [γ -³²P]ATP) at 25 °C for 30 min (for Trc) or 30 °C for 2 h (for Wts). The resulting solution was separated by SDS–PAGE and then analysed by PhosphorImager.

Received 17 May; accepted 25 July 2006.

Published online 13 August 2006.

- Wässle, H., Peichl, I. & Boycott, B. B. Dendritic territories of cat retinal ganglion cells. *Nature* **292**, 344–345 (1981).
- MacNeil, M. A. & Masland, R. H. Extreme diversity among amacrine cells: implication for function. *Neuron* **20**, 971–982 (1998).
- Jan, Y. N. & Jan, L. Y. The control of dendrite development. *Neuron* **40**, 229–242 (2003).
- Grueber, W. B., Jan, L. Y. & Jan, Y. N. Tiling of the *Drosophila* epidermis by multidendritic sensory neurons. *Development* **129**, 2867–2878 (2002).
- Grueber, W. B. *et al.* Dendrites of distinct classes of *Drosophila* sensory neurons

show different capacities for homotypic repulsion. *Curr. Biol.* **13**, 618–626 (2003).

- Sugimura, K. *et al.* Distinct developmental modes and lesion-induced reactions of dendrites of two classes of *Drosophila* sensory neurons. *J. Neurosci.* **23**, 3752–3760 (2003).
- Emoto, K. *et al.* Control of dendritic branching and tiling by the Tricornered-kinase/Furry signaling pathway in *Drosophila* sensory neurons. *Cell* **119**, 245–256 (2004).
- Justice, R. W. *et al.* The *Drosophila* tumor suppressor gene warts encodes a homolog of human myotonic dystrophy kinase and is required for the control of cell shape and proliferation. *Genes Dev.* **9**, 534–546 (1995).
- Xu, T. *et al.* Identifying tumor suppressors in genetic mosaics: the *Drosophila* *lats* gene encodes a putative protein kinase. *Development* **121**, 1053–1063 (1995).
- Tamaskovic, R., Bichsel, S. J. & Hemmings, B. A. NDR family of AGC kinases—essential regulators of the cell cycle and morphogenesis. *FEBS Lett.* **546**, 73–80 (2003).
- Harvey, K. F., Pfeiffer, C. M. & Hariharan, I. The *Drosophila* Mst ortholog, Hippo, restricts growth and cell proliferation and promotes apoptosis. *Cell* **114**, 457–467 (2003).
- Wu, S., Huang, J., Dong, J. & Pan, D. *hippo* encodes a Ste-20 family protein kinase that restricts cell proliferation and promotes apoptosis in conjugation with *salvador* and *warts*. *Cell* **114**, 445–456 (2003).
- Pantalacci, S., Tapan, N. & Leopold, P. The Salvador partner Hippo promotes apoptosis and cell-cycle exit in *Drosophila*. *Nature Cell Biol.* **5**, 921–927 (2003).
- Udan, R. S. *et al.* Hippo promotes proliferation arrest and apoptosis in the Salvador/Warts pathway. *Nature Cell Biol.* **5**, 914–920 (2003).
- Jia, J. *et al.* The *Drosophila* Ste20 family kinase dMST functions as a tumor suppressor by restricting cell proliferation and promoting apoptosis. *Genes Dev.* **17**, 2514–2519 (2003).
- Scott, E. K. & Luo, L. How dendrites take their shape? *Nature Neurosci.* **4**, 359–365 (2001).
- Cline, H. T. Dendritic arbor development and synaptogenesis. *Curr. Opin. Neurosci.* **11**, 118–126 (2001).
- Kaufmann, W. E. & Moser, H. W. Dendrite abnormality in disorders associated with mental retardation. *Cereb. Cortex* **10**, 981–991 (2000).
- Becker, L. E., Armstrong, D. L. & Chan, F. Dendritic atrophy in children with Down's syndrome. *Ann. Neurol.* **20**, 520–526 (1986).
- Zallen, J. A. *et al.* Neuronal cell shape and neurite initiation are regulated by the Ndr kinase SAX-1, a member of the Orb6/COT-1/warts serine/threonine kinase family. *Mol. Biol. Cell* **11**, 3177–3190 (2000).
- Gallegos, M. E. & Bargmann, C. I. Mechanosensory neurite termination and tiling depends on SAX-2 and SAX-1 kinase. *Neuron* **44**, 239–249 (2004).
- St John, M. A. *et al.* Mice deficient of Lats1 develop soft-tissue sarcomas, ovarian tumours and pituitary dysfunction. *Nature Genet.* **21**, 182–186 (1999).
- Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).
- Ruan, W., Pang, P. & Rao, Y. The SH2/SH3 adaptor protein dock interacts with Ste20-like kinase misshapen in controlling growth cone motility. *Neuron* **24**, 595–605 (1999).
- Chan, E. H. Y. *et al.* The Ste20-like kinase Mst2 activates the human large tumor suppressor kinase Lats1. *Oncogene* **24**, 2076–2086 (2005).
- Mikeladze-Dvali, T. *et al.* The growth regulators warts/lats and melted interact in a bistable loop to specify opposite fates in *Drosophila* R8 photoreceptors. *Cell* **122**, 775–787 (2005).
- Huang, J. *et al.* The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the *Drosophila* homolog of Yap. *Cell* **122**, 421–434 (2005).
- Stegert, M. R., Hergovich, A., Tamaskovic, R., Bichsel, S. J. & Hemmings, B. A. Regulation of NDR protein kinase by hydrophobic motif phosphorylation mediated by the mammalian Ste20-like kinase MST3. *Mol. Cell Biol.* **25**, 11019–11029 (2005).
- Lai, Z. C. *et al.* Control of cell proliferation and apoptosis by Mob as tumor suppressor, Mats. *Cell* **120**, 675–685 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to I. Hariharan, D. Pan, G. Halder, J. Jiang, C. Desplan, Y. Rao, N. Tapon, Z. C. Lai and Bloomington *Drosophila* Stock Center for fly stocks and reagents. We would like to thank I. Hariharan for suggestions, and M. Kim for helping to make transgenic flies. This work is supported by an NIH grant to Y.-N.J. J.Z.P. is supported by a Ruth L. Kirschstein NRSA postdoctoral fellowship. Y.-N.J. and L.Y.J. are Investigators of the Howard Hughes Medical Institute.

Author Contributions K.E., J.Z.P. and Y.-N.J. conceived and designed the experiments. K.E. performed the experiments. K.E., J.Z.P., L.Y.J. and Y.-N.J. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.-N.J. (yuhnung.jan@ucsf.edu).

LETTERS

The pathological response to DNA damage does not contribute to p53-mediated tumour suppression

M. A. Christophorou¹, I. Ringshausen¹, A. J. Finch¹, L. Brown Swigart¹ & G. I. Evan¹

The p53 protein has a highly evolutionarily conserved role in metazoans as ‘guardian of the genome’, mediating cell-cycle arrest and apoptosis in response to genotoxic injury¹. In large, long-lived animals with substantial somatic regenerative capacity, such as vertebrates, p53 is an important tumour suppressor—an attribute thought to stem directly from its induction of death or arrest in mutant cells with damaged or unstable genomes. Chemotherapy and radiation exposure both induce widespread p53-dependent DNA damage. This triggers potentially lethal pathologies² that are generally deemed an unfortunate but unavoidable consequence of the role p53 has in tumour suppression. Here we show, using a mouse model in which p53 status can be reversibly switched *in vivo* between functional and inactive states³, that the p53-mediated pathological response to whole-body irradiation, a prototypical genotoxic carcinogen, is irrelevant for suppression of radiation-induced lymphoma. In contrast, delaying the restoration of p53 function until the acute radiation response has subsided abrogates all of the radiation-induced pathology yet preserves much of the protection from lymphoma. Such protection is absolutely dependent on p19^{ARF}—a tumour suppressor induced not by DNA damage, but by oncogenic disruption of the cell cycle.

p53-deficient mice show a marked resistance to apoptosis of ‘radiosensitive’ tissues following systemic genotoxic injury⁴ but have a greatly increased risk of cancer^{5,6}. In the homozygous p53ER^{TAM} knockin (p53^{KI/KI}) mouse, each copy of the endogenous p53 gene is replaced with one encoding the p53ER^{TAM} fusion protein, which is completely dependent on the ligand 4-hydroxytamoxifen (4-OHT) for all measurable p53 functions. Without 4-OHT, p53^{KI/KI} mice are functionally p53-deficient but are rapidly and reversibly toggled to p53 wild-type status following systemic 4-OHT administration³. It is important to note that the 4-OHT ligand does not, in itself, activate p53ER^{TAM} but, rather, renders it competent to become activated by appropriate upstream signals. We used p53^{KI/KI} mice to explore the cause-and-effect relationship between the acute pathological response to systemic genotoxic injury and suppression of the consequent tumours.

Gamma (γ)-radiation is a prototypical carcinogen that acts by directly inducing DNA damage, principally single- and double-strand breaks. Exposure of C57BL/6 mice with a wild-type p53 genotype (p53^{+/+}) to 2.5-Gy γ-radiation induces widespread p53-dependent cell death in radiosensitive tissues, but the mice survive and are significantly protected from radiation-induced cancer. In contrast, irradiation of p53-deficient mice (p53^{-/-}) induces negligible pathology but 100% of animals are dead from lymphoma by 24 weeks (ref. 7). To ascertain when p53 function is required to suppress radiation-induced lymphoma, p53^{KI/KI} mice were treated with 4-OHT for six days to reinstate p53 function and then exposed to 2.5-Gy whole-body γ-radiation (the “concurrent p53 restoration”

cohort). A second, control cohort was treated with vehicle control instead of 4-OHT (the “no p53 restoration” cohort). Irradiation at 2.5 Gy was chosen after careful titration as a radiation dose that induces a p53-dependent, but not a p53-independent, apoptotic response in radiosensitive tissues³. As shown previously, irradiation induces a profound p53-dependent DNA-damage response in tissues of p53-competent mice³, which precipitates widespread apoptosis of primary lymphoid organs and intestinal epithelium³ and sustained

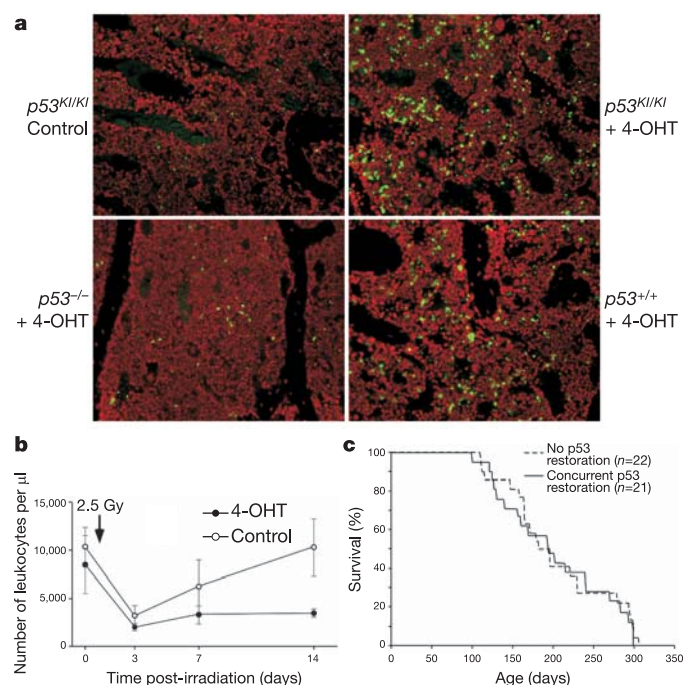


Figure 1 | The p53-dependent response to acute DNA damage induces multiple pathologies but no tumour suppression. **a**, TUNEL (TdT-mediated dUTP nick end labelling) assay for apoptosis of bone marrow from p53^{KI/KI} animals exposed to 2.5 Gy of whole-body γ-radiation in the presence of 4-OHT or vehicle (oil) control. p53^{+/+} and p53^{-/-} control animals treated with 4-OHT show that radiation-induced apoptosis is dependent on p53 function and not 4-OHT. Bone marrow was collected 4 h after irradiation. TUNEL-positive apoptotic cells are shown in green, with Hoechst nuclear counterstain in red. Original magnification, $\times 20$. **b**, Leukocyte counts of triplicate p53^{KI/KI} animals exposed to 2.5 Gy of γ-radiation in the presence of 4-OHT or vehicle control. Animals were irradiated on day 1. Error bars represent ± 1 s.d. **c**, Kaplan–Meier survival curves of p53^{KI/KI} mice in ‘concurrent restoration’ and ‘no restoration’ cohorts. Mice were treated with a single dose of 2.5 Gy of γ-radiation at 5–6 weeks of age.

¹Cancer Research Institute and Department of Cellular & Molecular Pharmacology, Comprehensive Cancer Center, University of California, San Francisco, California 94143, USA.

leukocytopenia (Fig. 1a, b). In contrast, control vehicle-treated $p53^{K1/K1}$ mice show negligible lymphoid or intestinal apoptosis³, only transient leukocytopenia (Fig. 1a, b), and no long-term tissue pathology, consistent with their uninterrupted p53-deficient status⁸. Thus, at this radiation dose functional p53 mediates all measurable radiation-induced pathologies in $p53^{K1/K1}$ mice. As whole-body irradiation is a potent carcinogen, both cohorts of mice were then monitored for tumour incidence. Despite their dramatic pathological response to acute irradiation, $p53^{K1/K1}$ mice in which p53 was functional concurrently with irradiation enjoyed no protection from early lymphoma incidence compared to control animals in which p53 function had never been restored (Fig. 1c).

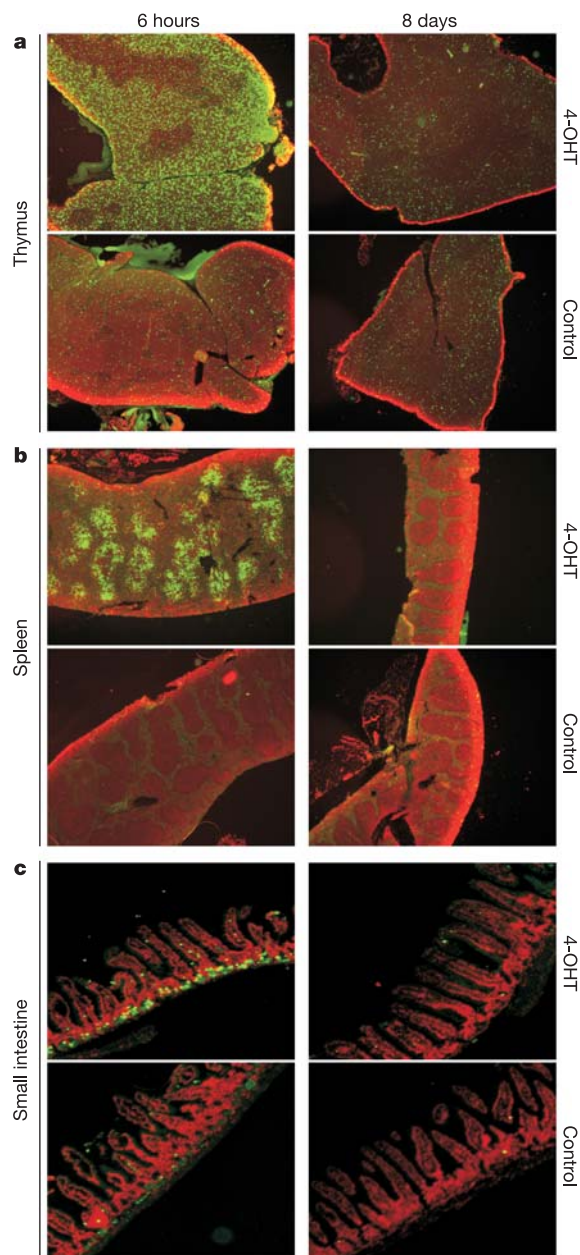


Figure 2 | Delaying restoration of p53 function abrogates radiation pathology. **a–c**, TUNEL assay of apoptosis in paraffin-embedded sections of thymus (**a**), spleen (**b**) and small intestinal epithelium (**c**) from $p53^{K1/K1}$ animals exposed to 2.5 Gy of γ -radiation at 5–6 weeks of age. p53 function was restored by systemic 4-OHT administration either 6 h or 8 days after irradiation. Tissues were collected 5 h after intraperitoneal administration of 4-OHT. TUNEL-positive apoptotic cells are shown in green, with Hoechst nuclear counterstain in red. Original magnification, $\times 2.5$.

Following acute, single-dose radiation exposure, global DNA damage in tissues is rapidly resolved, as indicated by the disappearance of DNA-damage-related markers such as phosphorylated γ H2AX (Supplementary Fig. 1) and phosphorylated ATM (data not shown). p53-activating signals decay with an identical kinetic and are undetectable after 48–72 h. The decay rate is identical in $p53^{K1/K1}$ and $p53^{+/+}$ animals (ref. 3 and Supplementary Fig. 2), indicating that p53ER^{TAM} is as effective as wild-type p53 in resolving acute DNA damage. Consistent with this rapid decay, restoration of p53 function eight days after irradiation induces no detectable apoptosis in radiosensitive tissues (Fig. 2) nor any pathology. To test whether restoration of p53 competence at this delayed time suppresses the emergence of radiation-induced lymphoma, $p53^{K1/K1}$ mice were exposed to 2.5-Gy γ -irradiation in the absence of 4-OHT, and p53 function was then restored eight days later and maintained for six days. Thus, the total period of p53 function in this “delayed restoration” cohort was identical to that in the “concurrent restoration” cohort. A schematic of these different regimens is shown in Fig. 3a, and Kaplan–Meier survival curves for the “no p53 restoration” versus the “delayed p53 restoration” cohorts in Fig. 3b. Delayed restoration of p53 offered significant protection ($P = 0.0007$) from radiation-induced lymphomagenesis, with a median delay in death of 99 days (see also Supplementary Fig. 3).

Together, these data demonstrate that—despite culling a large number of cells within the DNA-damaged populations from which subsequent lymphomas arise—the acute, p53-dependent pathology induced by whole-body genotoxic injury has a negligible role in the suppression of the eventual tumours induced by that insult. In contrast, delaying the restoration of p53 until after the global, acute DNA-damage response has subsided abrogates all radiation-induced pathology yet remains markedly tumour suppressive. Of note, only a brief period of transient p53 functional restoration is sufficient for substantial (~ 100 -day) delay in tumour onset.

The effectiveness of delayed p53 restoration in suppressing lymphomagenesis indicates that long after the p53-activating DNA-damage response has decayed in the bulk of cells, persistent p53-activating signals must remain in the small number of progenitor lymphoma cells engendered by the initial radiation insult.

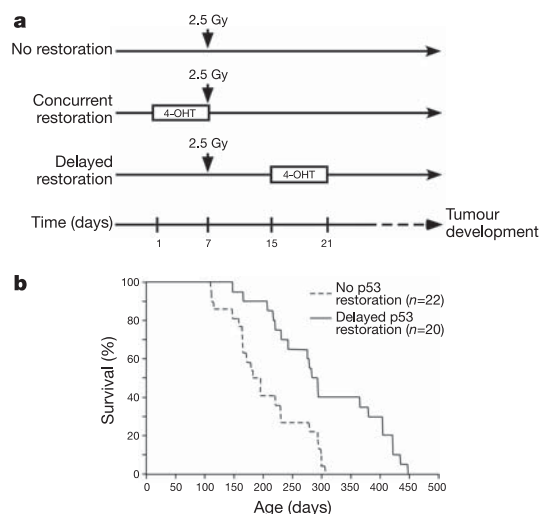


Figure 3 | Delaying restoration of p53 function until after the response to acute DNA damage has attenuated elicits marked tumour suppression.

a, Schematic of experimental design for tumorigenesis experiments. **b**, Kaplan–Meier analysis for the survival of $p53^{K1/K1}$ mice in the ‘delayed restoration’ cohort compared with the ‘no restoration’ cohort. Mice were treated with a single dose of 2.5 Gy of γ -radiation at 5–6 weeks of age. Log-rank analysis determined that the ‘delayed restoration’ cohort shows a significant delay in tumour development compared with the ‘no restoration’ cohort ($P = 0.0007$).

Recently, it was reported that early human cancerous lesions show evidence of persistent DNA damage^{9,10}, possibly as a result of telomere erosion or aberrant replication structures triggered by uncontrolled proliferation. Although telomere erosion is not extensive in inbred mouse tissues or tumours^{11,12}, and so cannot be the trigger for the delayed p53 activation we see, it remains possible that radiation injury generates unresolvable DNA damage that persistently signals to p53. Alternatively, p53 is also potentially^{13–17} and persistently³ activated by oncogenic mutations that disrupt cell-cycle control (for example, activation of Myc, Ras (Hras1) or E2F, or loss of Rb1), via induction of the tumour suppressor p19^{ARF} (refs 18–23). Notably, p19^{ARF} is not induced by DNA damage, and activates p53 directly, in the absence of any ancillary DNA damage or stress signals²⁴. As oncogene activation is likely to be an obligate attribute of incipient lymphoma cells, p19^{ARF} is a plausible trigger of p53 activity in such cells. Therefore, we tested directly whether p19^{ARF} is required for the tumour surveillance mediated by delayed p53 restoration. p19^{ARF}-null, p53^{K1/K1} (p53^{K1/K1};ARF^{-/-}) mice were irradiated as described above, and eight days later p53 function was reinstated for six days; the mice were then monitored for lymphoma.

As shown in Fig. 4a, delayed p53 restoration affords no protection from radiation-induced tumours if p19^{ARF} is absent. This demonstrates that p19^{ARF} is the crucial conduit of p53-mediated tumour suppression in incipient cancer cells, and also indicates that any persistent DNA damage signals—if they exist—are insufficient to activate p53 in the absence of p19^{ARF}. To confirm that p19^{ARF} status exerts no significant impact on DNA-damage responses, as published²⁴, p53 function was restored in p53^{K1/K1};ARF^{-/-} and p53^{K1/K1};ARF^{+/+} mice and the animals exposed to 2.5 Gy of γ -irradiation. Radiation induced robust p53-dependent target gene (*Cdkn1a* and *PUMA* (also known as *Bbc3*)) expression and cellular apoptosis in all radiosensitive tissues of p53^{K1/K1};ARF^{-/-} mice (Fig. 4b, c), and to similar extents irrespective of p19^{ARF} status. Non-4-OHT-treated control animals confirmed that the observed DNA-damage responses were p53-dependent. The pivotal role of p19^{ARF} as a persistent p53 trigger during tumour cell evolution is further reinforced by an analysis of the lymphomas that eventually arise in all of the p53^{K1/K1} cohorts. p53^{ER}^{TAM} is spontaneously activated on administration of 4-OHT to tumours that express p19^{ARF} but not in those that do not (Supplementary Fig. 4). Thus, even in established tumour cells, despite all of the additional lesions and potential genomic damage they may have accumulated, we see no evidence for any p53-activating signals other than p19^{ARF}.

In summary, our observations illustrate two key points. First, the acute pathological response to systemic genotoxic injury is irrelevant to suppression of eventual tumours induced by that injury. Second, restoring p53 function after the acute DNA-damage response has subsided engages substantial tumour suppression that is dependent on the p19^{ARF} pathway—a pathway activated not by DNA damage but by aberrant cell proliferation²⁵. These data raise the possibility that p53 sits at the unhappy confluence of two signals generated by DNA damage. Directly, DNA damage triggers widespread and transient p53 activation that is responsible for much of the pathology associated with systemic genotoxic injury, including extensive apoptosis of radiosensitive tissues. However, this widespread DNA damage contributes negligibly to tumour suppression. Indirectly, that same DNA damage triggers rare oncogenic mutations in a few surviving or misrepaired cells that, as a consequence, induce p19^{ARF} and elicit sustained p53 activation. It is this non-pathological, indirect pathway that seems to be crucial for the suppression of DNA-damage-induced lymphomagenesis (Supplementary Fig. 5).

Recently, several studies have called into question the widely held view that the DNA-damage response is integral to the actions of p53 as a tumour suppressor. Thus, replacement of endogenous p53 by a mutant that cannot be phosphorylated at the key Ser 18 residue (the mouse homologue of human p53 Ser 15) by DNA-damage-transducing kinases (Atm, Atr or Chk2) generates mice that are incompetent for DNA-damage-induced apoptosis yet fully protected from cancers²⁶. Moreover, absence of Atm has no impact on the ability of p53 to suppress tumours that are caused by cell-cycle-deregulating oncogenes in a transgenic brain tumour model²⁷. Notably, these investigations and our own are all *in vivo* studies, which skirt the complication of stress-induced p53 activation that confounds analyses of cells *in vitro*. It is also noteworthy that the close p53 relative p73 is potentially activated by DNA damage and shares most of the target genes and pro-apoptotic properties of p53²⁸. Unlike p53, however, p73 is not activated by p19^{ARF} and is an unconvincing tumour suppressor²⁹.

The possibility that the acute DNA damage response may be dispensable for p53-mediated tumour suppression has intriguing implications. First, although DNA damage is very efficient at triggering p53-dependent cell death, as evident from the widespread apoptosis of radiosensitive tissues following irradiation, it is highly inefficient at generating tumour cells: mice lacking p53 are immune to much of the pathological cell death induced by irradiation yet still take months to develop rare, clonal tumours. Thus, only a tiny number of damaged cells that p53 kills would ever have evolved into tumours had they survived. Widespread activation of p53 following

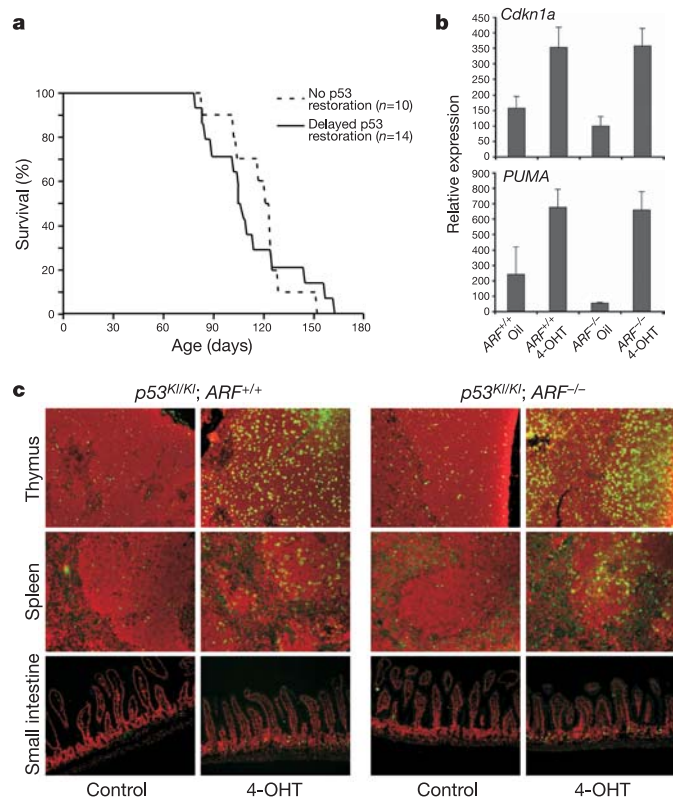


Figure 4 | p19^{ARF} mediates p53-dependent suppression of radiation-induced lymphoma but not p53-dependent DNA-damage responses.

a, Kaplan–Meier survival plots of p53^{K1/K1};ARF^{-/-} mice exposed to a single dose of 2.5 Gy whole-body γ -radiation at 5–6 weeks of age in the absence of p53 function ('no restoration') or the presence of p53 function but after the acute response to DNA damage has ceased ('delayed restoration').

b, Real-time quantitative polymerase chain reaction (PCR) analysis of total thymic RNA from p53^{K1/K1};ARF^{+/+} and p53^{K1/K1};ARF^{-/-} mice exposed to 2.5 Gy of γ -radiation in the presence of 4-OHT or vehicle control. Thymic RNA was harvested 5 h after irradiation. Expression of *Cdkn1a* and *PUMA* is shown relative to that of β -glucuronidase. Error bars represent ± 1 s.d. for triplicate mice for each condition and genotype. **c**, TUNEL staining for apoptosis in radiosensitive tissues of p53^{K1/K1};ARF^{-/-} mice exposed to 2.5 Gy of γ -radiation in the presence of 4-OHT or vehicle control. p53^{K1/K1};ARF^{+/+} animals are included as controls. Tissues were harvested 5 h after irradiation. TUNEL-positive apoptotic cells are shown in green, with Hoechst nuclear counterstain in red. Original magnification, $\times 20$.

DNA damage is therefore, at best, an unwieldy tool with which to cull a few potential tumour cells. In contrast, because p19^{ARF} is induced only in those rare cells that, as a consequence of DNA damage, acquire oncogenic mutations²⁵, its expression is highly specific to those few cells set on a neoplastic trajectory. Second, the p53-dependent DNA-damage response mediates much of the life-threatening pathology and side effects that accompany radiation exposure and chemotherapy. Restoring p53 function only at a later time abrogates all of the pathology yet, by focusing p53 activity only on those incipient tumour cells expressing p19^{ARF}, preserves much of the tumour suppression. Such observations suggest that transient pharmacological inhibition of p53 during, or shortly after, acute genotoxic injury may be helpful in ameliorating the pathology, as suggested², without compromising subsequent tumour suppression. Perhaps, most perversely, it is even possible that the linkage between p53 and DNA damage actually drives erosion of the efficacy of p53 as a tumour suppressor by imposing a lifelong selective pressure to inactivate p53 in cells sustaining such damage. Once lost, p53 is then unavailable to act as a tumour suppressor. Clearly, the extent to which the various attributes of p53—good and bad—can be separated by adroit manipulation is an area worthy of further investigation.

METHODS

Mouse breeding and handling. p53^{K1/K1} mice were bred and genotyped as described previously³. p53 function was restored by daily intraperitoneal injection of Tamoxifen in peanut oil vehicle. Where appropriate, 5–6-week-old mice were exposed to 2.5-Gy ionizing radiation from a ¹³⁷Cs source. Irradiated mice were monitored thrice-weekly for general health, well-being and tumour formation.

Immunohistochemistry and tissue apoptosis assays. Tissues and tumours were harvested and fixed in zinc-buffered formalin, paraffin embedded, sectioned, and stained with either haematoxylin and eosin or appropriate antibodies for classification. Blood was collected from the orbital sinus or plexus and haematology profiles were determined immediately using a Hemavet 850 machine (Drew Scientific).

Real-time quantitative PCR analysis. RNA was extracted with a Qiagen RNeasy Mini kit. Assay of RNA concentration and quality, DNase treatments, reverse transcription, and quantification were as described previously³.

Immunoblot analysis. Immunoblot analysis was as described previously³. p19^{ARF} protein was detected with rat monoclonal anti-p19^{ARF} primary antibody (clone C3, Novus) and β -actin with a mouse monoclonal anti- β -actin antibody (clone AC-15, Sigma).

Received 1 May; accepted 13 July 2006.

Published online 6 September 2006.

- Lane, D. P. p53, guardian of the genome. *Nature* **358**, 15–16 (1992).
- Gudkov, A. V. & Komarova, E. A. The role of p53 in determining sensitivity to radiotherapy. *Nature Rev. Cancer* **3**, 117–129 (2003).
- Christophorou, M. A. *et al.* Temporal dissection of p53 function *in vitro* and *in vivo*. *Nature Genet.* **37**, 718–726 (2005).
- Lain, S. & Lane, D. Improving cancer therapy by non-genotoxic activation of p53. *Eur. J. Cancer* **39**, 1053–1060 (2003).
- Jacks, T. *et al.* Tumor spectrum analysis in p53-mutant mice. *Curr. Biol.* **4**, 1–7 (1994).
- Donehower, L. A. *et al.* Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215–221 (1992).
- Kemp, C. J., Wheldon, T. & Balmain, A. p53-deficient mice are extremely susceptible to radiation-induced tumorigenesis. *Nature Genet.* **8**, 66–69 (1994).
- Komarova, E. A. & Gudkov, A. V. Could p53 be a target for therapeutic suppression? *Semin. Cancer Biol.* **8**, 389–400 (1998).
- Gorgoulis, V. G. *et al.* Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907–913 (2005).
- Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434**, 864–870 (2005).
- Prowse, K. R. & Greider, C. W. Developmental and tissue-specific regulation of mouse telomerase and telomere length. *Proc. Natl Acad. Sci. USA* **92**, 4818–4822 (1995).
- Kipling, D. & Cooke, H. J. Hypervariable ultra-long telomeres in mice. *Nature* **347**, 400–402 (1990).
- Wu, X. & Levine, A. J. p53 and E2F-1 cooperate to mediate apoptosis. *Proc. Natl Acad. Sci. USA* **91**, 3602–3606 (1994).
- Wagner, A. J., Kokontis, J. M. & Hay, N. Myc-mediated apoptosis requires wild-type p53 in a manner independent of cell cycle arrest and the ability of p53 to induce p21waf1/cip1. *Genes Dev.* **8**, 2817–2830 (1994).
- Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic *ras* provokes premature cell senescence associated with accumulation of p53 and p16^{INK4a}. *Cell* **88**, 593–602 (1997).
- Lowe, S. W., Cepero, E. & Evan, G. Intrinsic tumour suppression. *Nature* **432**, 307–315 (2004).
- Hermeking, H. & Eick, D. Mediation of c-Myc-induced apoptosis by p53. *Science* **265**, 2091–2093 (1994).
- Zindy, F. *et al.* Myc signaling via the ARF tumor suppressor regulates p53-dependent apoptosis and immortalization. *Genes Dev.* **12**, 2424–2433 (1998).
- Palmero, I., Pantoja, C. & Serrano, M. p19^{ARF} links the tumour suppressor p53 to Ras. *Nature* **395**, 125–126 (1998).
- de Stanchina, E. *et al.* E1A signaling to p53 involves the p19^{ARF} tumor suppressor. *Genes Dev.* **12**, 2434–2442 (1998).
- Bates, S. *et al.* p14^{ARF} links the tumour suppressors RB and p53. *Nature* **395**, 124–125 (1998).
- Quelle, D. E., Zindy, F., Ashmun, R. A. & Sherr, C. J. Alternative reading frames of the *INK4a* tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. *Cell* **83**, 993–1000 (1995).
- Mao, L. *et al.* A novel p16INK4A transcript. *Cancer Res.* **55**, 2995–2997 (1995).
- Kamijo, T. *et al.* Tumor suppression at the mouse *INK4a* locus mediated by the alternative reading frame product p19^{ARF}. *Cell* **91**, 649–659 (1997).
- Zindy, F. *et al.* Arf tumor suppressor promoter monitors latent oncogenic signals *in vivo*. *Proc. Natl Acad. Sci. USA* **100**, 15930–15935 (2003).
- Sluss, H. K., Armata, H., Gallant, J. & Jones, S. N. Phosphorylation of serine 18 regulates distinct p53 functions in mice. *Mol. Cell. Biol.* **24**, 976–984 (2004).
- Liao, M. J., Yin, C., Barlow, C., Wynshaw-Boris, A. & van Dyke, T. Atm is dispensable for p53 apoptosis and tumor suppression triggered by cell cycle dysfunction. *Mol. Cell. Biol.* **19**, 3095–3102 (1999).
- Harms, K., Nozell, S. & Chen, X. The common and distinct target genes of the p53 family transcription factors. *Cell. Mol. Life Sci.* **61**, 822–842 (2004).
- Soengas, M. S. & Lowe, S. W. p53 and p73: seeing double? *Nature Genet.* **26**, 391–392 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F. Rostker and the Genome and Pathology Cores at UCSF for technical assistance, J. Hwang for statistical analysis, J. H. Mao, F. McCormick, C. O'Shea, K. Shannon and the entire Evan laboratory for critical discussions. We thank T. Littlewood and D. Martin Zanca for showing that the p53ER^{1AM} knockin mouse was feasible. This work was supported by grants from the National Cancer Institute and National Institute on Aging.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.I.E. (gevan@cc.ucsf.edu).

LETTERS

Structure of the insulin receptor ectodomain reveals a folded-over conformation

Neil M. McKern¹, Michael C. Lawrence¹, Victor A. Streltsov¹, Mei-Zhen Lou¹, Timothy E. Adams¹, George O. Lovrecz¹, Thomas C. Elleman¹, Kim M. Richards¹, John D. Bentley¹, Patricia A. Pilling¹, Peter A. Hoyne¹, Kellie A. Cartledge¹, Tam M. Pham¹, Jennifer L. Lewis¹, Sonia E. Sankovich¹, Violet Stoichevska¹, Elizabeth Da Silva¹, Christine P. Robinson¹, Maurice J. Frenkel¹, Lindsay G. Sparrow¹, Ross T. Fernley¹, V. Chandana Epa¹ & Colin W. Ward¹

The insulin receptor is a phylogenetically ancient tyrosine kinase receptor found in organisms as primitive as cnidarians and insects. In higher organisms it is essential for glucose homeostasis¹, whereas the closely related insulin-like growth factor receptor (IGF-1R) is involved in normal growth and development². The insulin receptor is expressed in two isoforms, IR-A and IR-B; the former also functions as a high-affinity receptor for IGF-II and is implicated, along with IGF-1R, in malignant transformation³. Here we present the crystal structure at 3.8 Å resolution of the IR-A ectodomain dimer, complexed with four Fabs from the monoclonal antibodies 83-7 and 83-14 (ref. 4), grown in the presence of a fragment of an insulin mimetic peptide⁵. The structure reveals the domain arrangement in the disulphide-linked ectodomain dimer, showing that the insulin receptor adopts a folded-over conformation that places the ligand-binding regions in juxtaposition. This arrangement is very different from previous models⁶. It shows that the two L1 domains are on opposite sides of the dimer, too far apart to allow insulin to bind both L1 domains simultaneously as previously proposed⁷. Instead, the structure implicates the carboxy-terminal surface of the first fibronectin type III domain as the second binding site involved in high-affinity binding.

The structural organization of the insulin receptor (IR; Supplementary Fig. S1) is as follows. Each ectodomain monomer contains a leucine-rich repeat (L1) domain, a cysteine-rich (CR) region and a second leucine-rich repeat (L2) domain, followed by three fibronectin type III domains (FnIII-1 to FnIII-3). FnIII-2 contains an insert domain (ID) of 120 residues, within which lies the α-β cleavage site. The disulphide bond between each α- and β-chain involves the cysteines C647 and C860 (ref. 8). In addition, there are α-α disulphide bonds at C524 in the FnIII-1 domains and between the triplet C682, C683 and C685 in the ID⁸.

The crystallographic asymmetric unit of our structure comprises one ectodomain α-β chain pair (that is, the IRΔβ monomer), one 83-7 Fab and one 83-14 Fab. The IRΔβ dimer structure is shown in Fig. 1 and the comparative sequences and secondary structure assignments are shown in Supplementary Fig. S2. Each α-β monomer has an inverted 'V' layout with respect to the cell membrane (Fig. 1 and Supplementary Fig. S3). One leg of the V is formed by the L1, CR and L2 domains; the other is formed by an extended linear arrangement of the three FnIII domains. The ID is mostly disordered, consistent with DisProt⁹ predictions (data not shown). At the apex of the inverted V lies the connection between L2 and FnIII-1; these domains are not in extensive contact in the monomer (516 Å² buried molecular surface per interacting pair of L2 and FnIII-1 domains).

The arrangement of the L1, CR and L2 domains is similar, although not identical, to that observed for the human IR¹⁰ and IGF1-R¹¹ L1-CR-L2 fragment structures.

The IRΔβ monomer is paired in the unit cell to form the disulphide-bonded IR ectodomain homodimer (Fig. 1), which arises from a two-fold rotation of the monomer about an axis running roughly parallel to, but displaced ~10 Å from, the axis of the inverted V. The L2 domain of the first monomer contacts the FnIII-1 domain of the second monomer at the apex of the inverted V, whereas the L1 domain of the first monomer contacts the FnIII-2 domain of the second at the approximate midpoint of one of the two legs of the inverted V. The latter intermonomer interaction (878 Å² buried molecular surface per interacting pair of L1 and FnIII-2 domains),

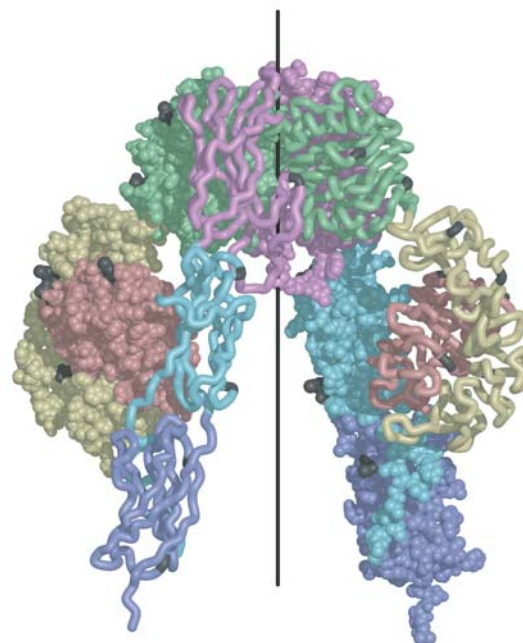


Figure 1 | The IRΔβ ectodomain homodimer, showing the juxtaposition of domains between the monomers. One monomer is shown in tube representation, the other in atomic sphere representation. Individual domains are coloured as follows: L1, brown; CR, yellow; L2, green; FnIII-1, cyan; FnIII-2, magenta; and FnIII-3, blue. The locations of potential N-linked glycosylation sites are shown in black.

¹CSIRO Molecular & Health Technologies, 343 Royal Parade, Parkville, Victoria 3052, Australia.

although not extensive, is relatively well packed and contains two potential salt bridges (E124–R780 and R804–E153 or R804–E154). The other intermonomer interaction (L2 and FnIII-1, burying 1,344 Å² of molecular surface per interacting pair of domains) is more extensive, but less well-packed, and contains two potential salt bridges (R454–D496 and E394–R498). Both pairs of salt bridges are potentially conserved in IGF-1R (Supplementary Fig. S2). Located at the apex of the IRΔβ dimer, in close proximity to each other, is the pair of K460 residues in the last leucine-rich repeat in L2 (Fig. 2). These residues have been implicated both in regulating cooperative interactions between the two monomers in the IR dimer and as the sites for chemical crosslinking with disuccinimidyl suberate¹².

The C termini of the FnIII-3 domains are at the base of the inverted V structure where, in the intact IR, they would extend through the cell membrane (Supplementary Fig. S1). The N-linked glycosylation sites in the L1–CR–L2 fragment are distributed over its outer surface within the homodimer (Fig. 1), away from the surfaces involved in monomer–monomer interactions or in ligand binding (see below). By contrast, the glycosylated asparagines in the FnIII domains are towards the inner surface of the FnIII domain leg (Fig. 1 and Supplementary Fig. S3).

The structure provides experimental confirmation of three FnIII domains in the IR (see Supplementary Figs S1 and S2). Although previous predictions of the FnIII boundaries and strand locations are seen to be largely correct², none positioned the C' strands of FnIII-1 or FnIII-2 correctly or placed the ID within the CC' loop of FnIII-2, and only one of the predictions in each case was correct for the E strands of FnIII-1 (ref. 13) or FnIII-3 (ref. 14).

The inter-α-chain disulphide bond involving the pair of C524

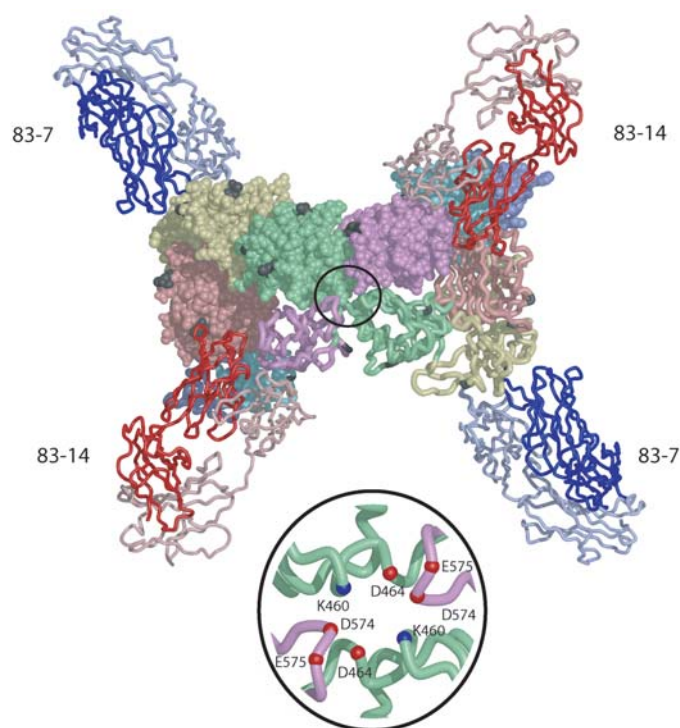


Figure 2 | The IRΔβ ectodomain homodimer, showing the attached pairs of Fabs. The IRΔβ homodimer is depicted with one molecule in tube representation and the other in atomic sphere representation. Domains are coloured as in Fig. 1 and the view is down the two-fold axis from the membrane distal end. The Fab fragments are represented as thin tubes and coloured as follows: 83-7 heavy chain, dark blue; 83-7 light chain, light blue; 83-14 heavy chain, red; and 83-14 light chain, pink. Glycosylation sites are shown in black. The inset shows the location of the pairs of charged residues K460, D464, D574 and E575 at the apex of the homodimer. The sequences of Fabs 83-7 and 83-14 are given in Supplementary Figs S5 and S6.

residues⁸ lies in a large, 28-residue loop (508–535) linking the C and C' strands of the FnIII-1 domain. The α–β disulphide involving C647–C860 can be observed: C647 is in the ID in a long extension of the C strand of FnIII-2 that runs parallel to FnIII-3 and extends down on the inside of the inverted V, almost to the bottom of FnIII-3, with the pair of S655 residues at the end of the visible electron density. Its partner, C860, is located close to the C terminus of the C' strand of FnIII-3. This additional connection between FnIII-2 and FnIII-3 would reduce the flexibility at the junction of these two domains. Such rigidity may be important in signal transduction because removal of this α–β disulphide bond by the C647S mutation abolishes basal and insulin-stimulated phosphorylation without affecting insulin binding¹⁵. An internal disulphide bridge links C786 and C795 at the base of the FG loop in FnIII-2. The unpaired C872 has its side chain partially buried in FnIII-3 and has no counterpart in the IGF-1R (Supplementary Fig. S2) or the IR-related receptor (IRR)⁸.

Low-resolution simulated projection images generated from the IRΔβ structure, both with and without the 83-7 and 83-14 Fabs attached, can be compared with experimental negative stain electron micrograph images obtained previously¹⁶ (Fig. 3). The correspondence of the projected structure with these images is marked, suggesting that the quaternary structure of the ectodomain observed in the crystal is similar to that of the particles observed in the electron microscopy studies¹⁶. The epitope for 83-7 Fab, which does not inhibit insulin binding⁴, is located in the segment 233–281 of the CR domain, whereas the monoclonal antibody (Mab) 83-14, which does inhibit insulin binding to IR on cells⁴, binds only to FnIII-1 (Fig. 2). The four Fabs protrude as extended 'handles' from the IR ectodomain homodimer (Figs 2 and 3), with their long axes lying at ~65° (Fab 83-14) and ~120° (Fab 83-7) with respect to the two-fold axis of the homodimer.

We are unable to correlate our structure with that proposed for the whole receptor in complex with insulin based on a three-dimensional reconstruction from scanning transmission electron microscopy dark-field images⁷. In that model, the L1 domains protrude at the membrane distal end of the molecule and the FnIII-2 and FnIII-3 domains lie coplanar to the membrane⁷. By contrast, our crystal structure shows that the L1 domains lie towards the centre of the molecule at the start of one leg of the inverted V monomer structure with the FnIII domains arranged in a linear manner forming the other leg (Fig. 1 and Supplementary Fig. S3). The ligand-binding faces of the two L1 domains are thus located on opposite sides of the

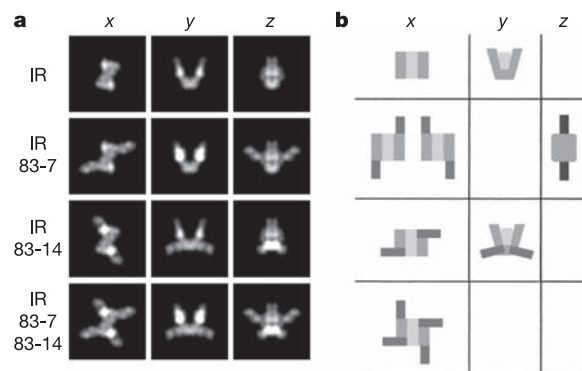


Figure 3 | Comparison of the crystal structure of the Fab-complexed IRΔβ ectodomain with single-particle electron microscope images. **a**, Simulated projection images at ~20 Å resolution obtained from the crystal structure presented here (see Supplementary Methods). **b**, Proposed projected views of previously observed particles (reprinted, with permission, from ref. 16). The correspondence between **a** and **b** is marked. (The z-projection in **b** of IRΔβ complexed with 83-7 Fab does not correspond to its partner view in **a** and is probably a variant of the x-projection of the same complex.)

Our structure suggests that this second region of insulin binds to one or more of the three loops (AB, CC' and EF) in FnIII-1 of

The formation of the high-affinity state is associated with a rearrangement of IR domains^{18,19,24}, which in turn leads to transphosphorylation of the intracellular kinase and regulatory regions of the receptor. The details of these domain movements remain to be established because soluble IR ectodomain does not undergo ligand-induced transitions unless it is tethered at its C-terminal end³⁰. Given the extensive interaction between the L1 and CR domains in each monomer and the expected rigidity of the FnIII-2–FnIII-3 connections as a result of the α – β disulphide bonds, the points at which these domain movements potentially occur in the monomer are the CR–L2, L2–FnIII-1 and FnIII-1–FnIII-2 junctions. The interface between the L1 domain of one monomer and the FnIII-2 domain of the other suggests that any closure of the L1–CR unit of one monomer towards the C-terminal end of the FnIII-1 domain of the other would also drag the FnIII-2–FnIII-3 unit with it, thereby changing the relative position of the intracellular kinase domains in the IR dimer and potentially inducing signalling (see Supplementary Information). Such arguments remain speculative and high-resolution structures of insulin–IR complexes will be required to establish the details of ligand–receptor binding and the domain rearrangements that accompany signalling.

Phasing, model building and refinement. Experimental details, references and final refinement statistics are given in the Supplementary Methods and Supplementary Table S1. The locations of the L1, CR, L2 and Fab domains were found by molecular replacement using PHASER with native data set 1. Search models were the L1-CR and the L2 domains from the L1-CR-L2 fragment¹⁰ and the Hy-HEL5 antibody (PDB entry 1BQL) set at 5° hinge angle increments. Phase improvement was obtained with RESOLVE and BUSTER-TNT, and the resultant electron density map showed the likely location of the three FnIII like domains. The molecular replacement solution was confirmed using native data set 2 including, in the process, phase information derived from the PIP-derivative data set (Supplementary Table S1), processed with SHARP and subjected to solvent-flattening with DM and/or SOLOMON. Model building of the FnIII domains using domains 7, 8 and 9 of the fibronectin structure (PDB code 1FNF) as a guide, crystallographic refinement using iterative cycles of BUSTER-TNT and/or REFMAC5, and manual model building using XtalView/Xfit are described in the Supplementary Methods. Within BUSTER-TNT, scattering from the missing atoms was modelled with a low-resolution homographic exponential distribution, and maximum entropy density completion was used at the end of each round of refinement to recover the density for missing parts of the structure.

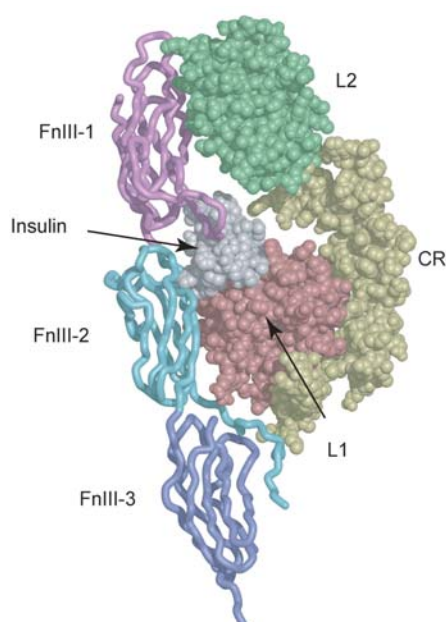


Figure 4 | Insulin bridging two monomers in the IRΔβ homodimer. The L1, CR and L2 domains (atomic sphere representation) are from one monomer in the IR homodimer, whereas the FnIII domains (tube representation) are from the other. Domains are represented and coloured as in Fig. 1. Insulin (grey atomic sphere representation) is positioned on L1 according to the model of insulin bound to the L1-CR-L2 fragment¹⁰ with no further adjustments.

Received 19 April; accepted 21 July 2006.

Published online 6 September 2006.

1. Kitamura, T., Kahn, C. R. & Accili, D. Insulin receptor knockout mice. *Annu. Rev. Physiol.* **65**, 313–332 (2003).
2. Adams, T. E., Epa, V. C., Garrett, T. P. J. & Ward, C. W. Structure and function of the type 1 insulin-like growth factor receptor. *Cell. Mol. Life Sci.* **57**, 1050–1093 (2000).
3. Denley, A., Wallace, J. C., Cosgrove, L. J. & Forbes, B. E. The insulin receptor isoform exon 11[−] (IR-A) in cancer and other diseases: a review. *Horm. Metab. Res.* **35**, 778–785 (2003).
4. Soos, M. A. *et al.* Monoclonal antibodies reacting with multiple epitopes on the human insulin receptor. *Biochem. J.* **235**, 199–208 (1986).
5. Schaffer, L. *et al.* Assembly of high-affinity insulin receptor agonists and antagonists from peptide building blocks. *Proc. Natl Acad. Sci. USA* **100**, 4435–4439 (2003).
6. De Meyts, P. & Whittaker, J. Structural biology of insulin and IGF1 receptors: implications for drug design. *Nature Rev. Drug Discov.* **1**, 769–783 (2002).
7. Ottensmeyer, F. P., Beniach, D. R., Luo, R. Z. T. & Yip, C. C. Mechanism of transmembrane signaling: insulin binding and the insulin receptor. *Biochemistry* **39**, 12103–12112 (2000). Corrigendum. *Ibid.* **40**, 6988.
8. Sparrow, L. G. *et al.* The disulfide bonds in the C-terminal domains of the human insulin receptor ectodomain. *J. Biol. Chem.* **272**, 29460–29467 (1997).
9. Peng, K. *et al.* Optimizing long intrinsic disorder predictors with protein evolutionary information. *J. Bioinform. Comput. Biol.* **3**, 35–60 (2005).
10. Lou, M. *et al.* Crystal structure of the first three domains of the human insulin receptor reveals major differences from the IGF-1 receptor in the regions governing ligand specificity. *Proc. Natl Acad. Sci. USA* **103**, 12429–12434 (2006).
11. Garrett, T. P. J. *et al.* Crystal structure of the first three domains of the type-1 insulin-like growth factor receptor. *Nature* **394**, 395–399 (1998).
12. Kadowaki, H. *et al.* Mutagenesis of lysine 460 in the human insulin receptor: effects upon receptor recycling and cooperative interactions among binding sites. *J. Biol. Chem.* **265**, 21285–21296 (1990).
13. Marino-Buslje, C., Martin-Martinez, M., Mizuguchi, K., Siddle, K. & Blundell, T. L. The insulin receptor: from protein sequence to structure. *Biochem. Soc. Trans.* **27**, 715–726 (1999).
14. Schaefer, E. M., Erickson, H. P., Federwisch, M., Wollmer, A. & Ellis, L. Structural organization of the human insulin receptor ectodomain. *J. Biol. Chem.* **267**, 23393–23402 (1992).
15. Cheatham, B. & Kahn, C. R. Cysteine 647 in the insulin receptor is required for normal covalent interaction between α - and β -subunits and signal transduction. *J. Biol. Chem.* **267**, 7108–7115 (1992).
16. Tulloch, P. A. *et al.* Single-molecule imaging of human insulin receptor ectodomain and its Fab complexes. *J. Struct. Biol.* **125**, 11–18 (1999).
17. Burgess, A. W. *et al.* An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol. Cell* **12**, 541–552 (2003).
18. Lee, J., Pilch, P. F., Shoelson, S. E. & Scarlata, S. F. Conformational changes of the insulin receptor upon insulin binding and activation as monitored by fluorescence spectroscopy. *Biochemistry* **36**, 2701–2708 (1997).
19. Florke, R. R. *et al.* Hormone-triggered conformational changes within the insulin-receptor ectodomain: requirement for transmembrane anchors. *Biochem. J.* **360**, 189–198 (2001).
20. Yip, C. C. & Ottensmeyer, P. Three-dimensional structural interactions of insulin and its receptor. *J. Biochem.* **278**, 27329–27332 (2003).
21. Surinya, K. H. *et al.* Role of insulin receptor dimerization domains in ligand binding, cooperativity, and modulation by anti-receptor antibodies. *J. Biol. Chem.* **277**, 16718–16725 (2002).
22. Kurose, T. *et al.* Cross-linking of a B25 azidophenylalanine insulin derivative to the carboxy-terminal region of the α -subunit of the insulin receptor. Identification of a new insulin-binding domain in the insulin receptor. *J. Biol. Chem.* **269**, 29190–29197 (1994).
23. Kristensen, C., Wiberg, F. C. & Andersen, A. S. Specificity of insulin and insulin-like growth factor I receptors investigated using chimeric mini-receptors—role of C-terminal of receptor α subunit. *J. Biol. Chem.* **274**, 37351–37356 (1999).
24. Schaffer, L. A model for insulin binding to the insulin receptor. *Eur. J. Biochem.* **221**, 1127–1132 (1994).
25. De Meyts, P. The structural basis of insulin and IGF-1 receptor binding and negative co-operativity, and its relevance to mitogenic versus metabolic signaling. *Diabetologia* **37** (suppl. 2), S135–S148 (1994).
26. De Meyts, P. Insulin and its receptor: structure, function and evolution. *BioEssays* **26**, 1351–1362 (2004).
27. Zhang, B. & Roth, R. A. A region of the insulin receptor important for ligand binding (residues 450–601) is recognized by patient's autoimmune antibodies and inhibitory monoclonal antibodies. *Proc. Natl Acad. Sci. USA* **88**, 9858–9862 (1991).
28. Schumacher, R. *et al.* Signaling-competent receptor chimeras allow mapping of major insulin receptor binding domain determinants. *J. Biol. Chem.* **268**, 1087–1094 (1993).
29. Fabry, M. *et al.* Detection of a new hormone contact site within the insulin receptor ectodomain by the use of a novel photoreactive insulin. *J. Biol. Chem.* **267**, 8950–8956 (1992).
30. Hoyne, P. A. *et al.* High affinity insulin binding by soluble insulin receptor extracellular domain fused to a leucine zipper. *FEBS Lett.* **479**, 15–18 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Siddle for the 83-7 and 83-14 monoclonal antibody cell lines; O. Dolezal for advice on cloning and sequencing the monoclonal antibody variable region cDNAs; B. Van Donkelaar for technical crystallography inputs; and L. Lu, L. Cheong and T. Phan for contributions to protein production. We acknowledge the help provided by beamline staff at both the Advanced Photon Source and the Photon Factory. This work was supported by the Australian Synchrotron Research Program, which is funded by the Commonwealth of Australia under the Major National Research Facilities Program. Additional financial support was provided under the Generic Technology component of the Industry Research and Development Act 1986, from Biota Diabetes Research Pty Ltd.

Author Information Atomic coordinates of the IR-Fab complex have been deposited in the Protein Data Bank with accession number 2DTG. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.W.W. (Colin.Ward@csiro.au) or M.C.L. (Mike.Lawrence@csiro.au).

LETTERS

Autophosphorylation at serine 1987 is dispensable for murine *Atm* activation *in vivo*

Manuela Pellegrini^{1*}, Arkady Celeste^{1*}, Simone Difilippantonio¹, Rong Guo², Weidong Wang², Lionel Feigenbaum³ & André Nussenzweig¹

The ATM (ataxia telangiectasia mutated) protein kinase is activated under physiological and pathological conditions that induce DNA double-strand breaks (DSBs). Loss of ATM or failure of its activation in humans and mice lead to defective cellular responses to DSBs, such as cell cycle checkpoints, radiation sensitivity, immune dysfunction, infertility and cancer predisposition. A widely used biological marker to identify the active form of ATM is the autophosphorylation of ATM at a single, conserved serine residue (Ser 1981 in humans; Ser 1987 in mouse)¹. Here we show that *Atm*-dependent responses are functional at the organismal and cellular level in mice that express a mutant form of *Atm* (mutation of Ser to Ala at position 1987) as their sole *Atm* species. Moreover, the mutant protein does not exhibit dominant-negative interfering activity when expressed physiologically or overexpressed in the context of *Atm* heterozygous mice. These results suggest an alternative mode for stimulation of *Atm* by DSBs in which *Atm* autophosphorylation at Ser 1987, like transphosphorylation of downstream substrates, is a consequence rather than a cause of *Atm* activation.

It has been demonstrated that ATM is sequestered as an inactive dimer or higher order multimer in the unperturbed cell¹. The generation of a DSB was proposed to induce a relaxation in the surrounding chromatin that promotes the intermolecular phosphorylation of ATM at Ser 1981 (ref. 1). This autophosphorylation releases inhibitory contacts between the catalytic domain in one ATM molecule and the region surrounding Ser 1981 in an interacting partner, resulting in the dissociation of ATM dimers into active monomers. After activation, phosphorylated ATM monomers are recruited to DNA breaks by binding to the carboxy terminus of the DSB sensor NBS1 (refs 2, 3), where they phosphorylate numerous substrates. This model predicts that a non-phosphorylatable mutant ATM protein would have dominant-negative interfering activity in cells expressing wild-type ATM, and would be incapable of reconstituting ATM-deficient cells. Both predictions have been confirmed in transformed human cell lines in which the mutant protein is ectopically expressed¹.

To determine the role of ATM autophosphorylation in a physiological context, we chose a bacterial artificial chromosome (BAC) reconstitution method for the generation of a mouse model. As a proof of principle, transgenic mice from two independent founder lines expressing a wild-type BAC (RP24-122F10) spanning the genomic mouse *Atm* locus (*Atm*^{Tg^{WT}}; Supplementary Fig. 1a) were backcrossed to *Atm*^{-/-} mice. Expression of the wild-type mouse transgene in the *Atm* null background rescued the defects in growth, lymphocyte development and meiotic arrest of the homozygous

mutant mice. By contrast, a BAC containing human ATM (RP11-455M10) failed to express in the mouse (not shown).

We then mutated the 1987 serine residue of *Atm* to alanine in the mouse BAC using an oligonucleotide-based recombination method⁴. The presence of S1987A was confirmed by sequence analysis (Supplementary Fig. 1b), and the mutated BAC was subsequently used to generate transgenic mice. Several founder lines were produced, two of which were analysed in detail. In the first (B1), mutant *Atm*-S1987A protein expression was similar to endogenous levels in *Atm*^{+/-} mice irrespective of the tissue (splenocytes, Fig. 1a, b; thymocytes and testes, not shown), whereas the second transgenic founder (A8) produced 12-fold higher levels of *Atm* compared to *Atm*^{+/-} mice

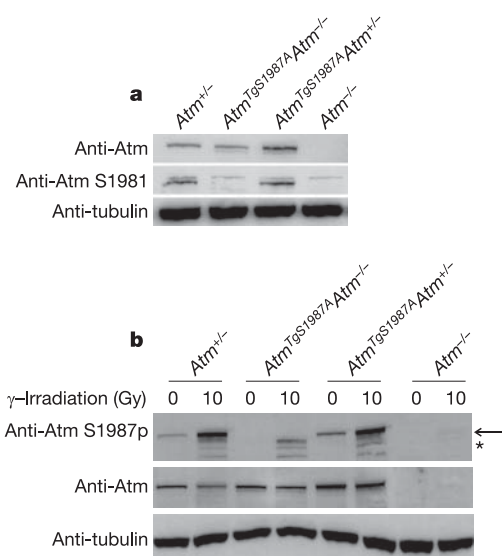


Figure 1 | *Atm* null and heterozygous mice express *Atm*-S1987A from a BAC transgene. **a, Unphosphorylated *Atm* S1981 and total *Atm* protein in B cells from transgene (line B1) positive and negative littermates. **b**, B cells from *Atm*^{+/-}, *Atm*^{-/-}, *Atm*^{TgS1987A}/*Atm*^{+/-} and *Atm*^{TgS1987A}/*Atm*^{-/-} mice (B1) were harvested 30 min after no treatment (0 Gy) or after γ-irradiation (10 Gy), and phosphorylated *Atm* at S1987 was assessed by immunoblotting. Similar results were found for transgenic line A8 that expresses 12-fold higher levels of *Atm*-S1987A (see Supplementary Fig. 1c, d). The asterisk corresponds to a cross-reactive protein, subject to *Atm*- and damage-dependent phosphorylation, that migrates faster than authentic phosphorylated *Atm* at S1987 (indicated by arrow).**

¹Experimental Immunology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-1360, USA. ²Laboratory of Genetics, National Institute on Aging, National Institutes of Health, Baltimore, Maryland 21224, USA. ³SAIC-Frederick, National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, Maryland 21702, USA.

*These authors contributed equally to this work.

(Supplementary Fig. 1c). Using an antibody that detects the unphosphorylated S1981 residue¹, we confirmed that *Atm*^{TgS1987A}*Atm*^{-/-} mice do not express *Atm* containing serine at residue 1987 (Fig. 1a). Moreover, autophosphorylation at S1987 was undetectable in *Atm*^{TgS1987A}*Atm*^{-/-} B cells exposed to irradiation, as assessed using phospho-specific monoclonal antibodies raised against mouse *Atm* S1987p (Fig. 1b and Supplementary Fig. 1d) or human *ATM* S1981p (Supplementary Fig. 2). We conclude that the mutant *Atm*-S1987A protein is the only potential source of *Atm* kinase activity in *Atm*^{TgS1987A}*Atm*^{-/-} mice.

Antigen receptor rearrangements in lymphocytes generate DSBs that have the potential to phosphorylate *Atm* at S1987 *in vivo*⁵. Consistent with the activation of *ATM* during physiological V(D)J and immunoglobulin class-switch recombination, loss of *ATM* results in defective lymphocyte development and function, and accumulation of chromosomal aberrations involving antigen receptor loci. Expression of *Atm*-S1987A in the *Atm*^{-/-} background rescued these defects, as demonstrated by the normal levels of mature single positive CD4⁺T cells and expression of TCR- β in *Atm*^{TgS1987A}*Atm*^{-/-} mice (Fig. 2a, b and Supplementary Fig. 3a). Similarly, B cells derived from *Atm*^{TgS1987A}*Atm*^{-/-} mice and activated to undergo class-switch recombination produced normal levels of surface IgG1 (Fig. 2c and Supplementary Fig. 3b). Moreover, they did not accumulate a high frequency of chromosomal aberrations, in stark contrast to stimulated *Atm*^{-/-} B cells⁶ (Fig. 2d and Supplementary Fig. 3b). Thus, *Atm*^{TgS1987A}*Atm*^{-/-} lymphocytes do not exhibit any of the hallmark defects found in *Atm*^{-/-} mice.

Although it is formally possible that *Atm* (and the mutant *Atm*-S1987A protein) functions in a structural (rather than a catalytic) capacity to facilitate the processing of physiological V(D)J and class-switch recombination-induced DSBs, *ATM* kinase activity is known

to be essential for cellular signalling in response to external damage. For example, *ATM*-dependent phosphorylation of Smc1, Chk2 and Chk1 contributes to the irradiation-induced intra-S-phase and G2/M checkpoints^{7–11}. To determine whether mutant *Atm*-S1987A protein is capable of triggering DNA-damage-induced cell cycle checkpoints, we treated primary B cells with γ -irradiation. *Atm*^{TgS1987A}*Atm*^{-/-} B cells assayed over a dose range of 0.5–10 Gy showed a reduction in the mitotic index and in the rate of DNA replication comparable to *Atm*^{+/+} and *Atm*^{+/-} controls, indicative of a normal G2/M and intra-S-phase cell cycle checkpoint arrest, which contrasted with cells from *Atm*^{-/-} littermates (Supplementary Figs 3c and 4a, b). Moreover, *Atm*^{TgS1987A}*Atm*^{-/-} and *Atm*^{TgS1987A}*Atm*^{+/-} ear fibroblasts were not hypersensitive to γ -irradiation (Supplementary Fig. 4c). To examine directly DNA damage signalling, we measured *Atm* substrate phosphorylation in primary B cells. γ -Irradiation-induced phosphorylation of Smc1, Chk2, Chk1 and p53 was barely detectable in *Atm*^{-/-} B cells 30 min after treatment with 10 Gy γ -irradiation (Fig. 3a and Supplementary Fig. 1d). In contrast, *Atm*^{TgS1987A}*Atm*^{-/-} B cells exhibited normal induction (Fig. 3a and Supplementary Fig. 1d) and kinetics (Supplementary Fig. 5a) of *Atm* substrate phosphorylation when compared to controls. Pre-treatment of *Atm*^{TgS1987A}*Atm*^{-/-} cells with a small molecule inhibitor of *Atm* (KU55933; Supplementary Fig. 5b)¹², but not a DNA-PKcs inhibitor (NU7026; Supplementary Fig. 5c), abrogated γ -irradiation-induced phosphorylation of Chk2 and Smc1 and reduced γ -H2AX formation. In contrast, KU55933 did not affect Atr-mediated phosphorylation of Chk1 in response to hydroxyurea (Supplementary Fig. 5d). Thus, *Atm* and not related kinases such as Atr and DNA-PK mediate the γ -irradiation-induced phosphorylation events in *Atm*^{TgS1987A}*Atm*^{-/-} mice. We therefore conclude that the mutant *Atm*-S1987A protein exhibits normal *Atm*

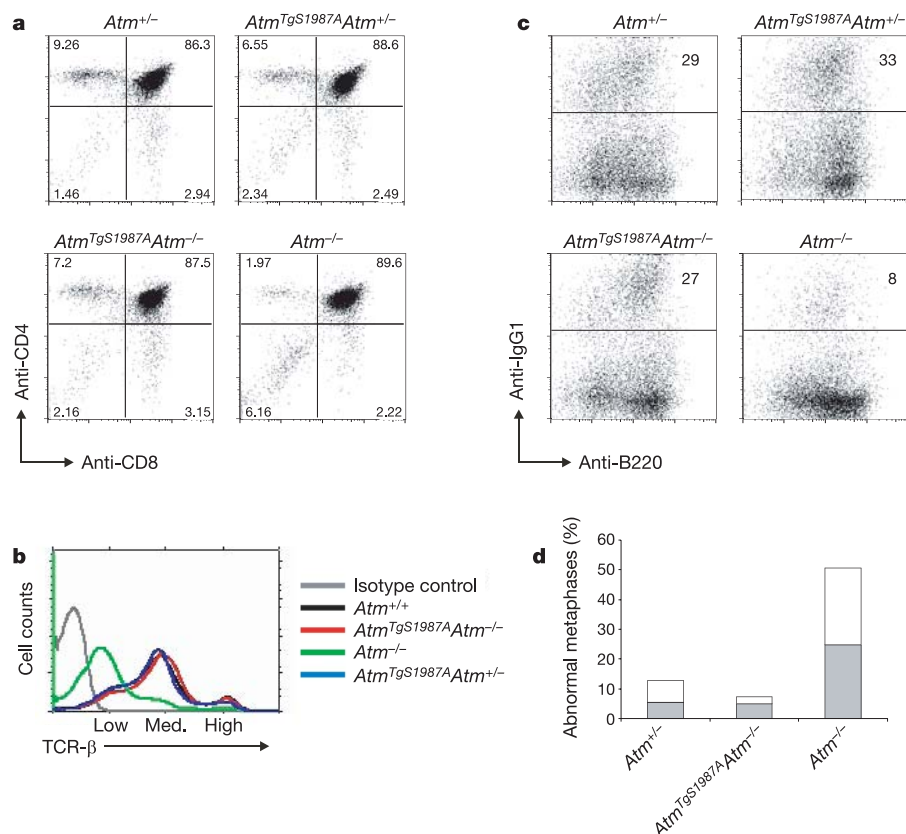


Figure 2 | Expression of *Atm*-S1987A reconstitutes lymphocyte development and restores genomic stability. **a**, Surface expression of CD4 versus CD8 in freshly isolated thymocytes. **b**, Surface expression of T-cell receptor- β (TCR- β). **c**, Efficiency of class switching to IgG1 in transgenic B

cells. **d**, Percentage of metaphases with abnormalities specifically associated with chromosome 12 (filled bar) and with all other chromosomes (open bar) in B cells stimulated with LPS plus IL-4 for 72 h.

kinase activity, which is sufficient to trigger an efficient checkpoint response to γ -irradiation.

Many ATM substrates are phosphorylated at sites of DSBs¹¹. To determine whether mutant *Atm*-S1987A is recruited to sites of DNA damage, we induced DSBs in primary fibroblasts and B cells using a laser microbeam, and the assembly of *Atm* coincident with γ -H2AX-marked sites of DSBs was tracked by fluorescence microscopy. We found that both wild-type and mutant *Atm* were rapidly recruited to breaks (Fig. 3b and Supplementary Fig. 6a). In addition, there was enhanced retention of *Atm*-S1987A at damaged sites in response to γ -irradiation, as demonstrated by the increase in protein that was resistant to detergent extraction in damaged *Atm*^{TgS1987A}*Atm*^{-/-} B cells (Supplementary Fig. 6b). Thus, the dynamics of *Atm*-S1987A re-localization to DSBs seems to be similar to wild-type *Atm*.

Ectopic expression of ATM-S1981A protein has been reported to have dominant interfering activity in transformed cells¹. However, our analysis of primary cells from *Atm*^{TgS1987A}*Atm*^{+/-} mice that expressed physiological levels of transgenic *Atm* (line B1) revealed that *Atm*-S1987A does not interfere with endogenous *Atm* autophosphorylation (Fig. 1b and Supplementary Fig. 2), trans-phosphorylation

of substrates (Fig. 3a), γ -irradiation-induced cell cycle checkpoints (Supplementary Fig. S4a, b) or chromatin binding (Supplementary Fig. 6b). Moreover, we excluded the possibility that overexpression of the *Atm*-S1987A protein contributes to a dominant interfering phenotype, as expression of 12-fold higher levels of *Atm*-S1987A in transgenic line A8 did not compromise endogenous *Atm* or downstream target phosphorylation (Supplementary Fig. 1d). Moreover, overexpressed *Atm*-S1987A was fully functional because A8-reconstituted *Atm*^{-/-} mice rescued defects in damage signalling, lymphocyte development, cell cycle checkpoints, meiosis and radiation sensitivity (Supplementary Figs 1d, 3 and 7). We conclude that the S1987A non-phosphorylatable protein does not interfere with wild-type *Atm*.

Autophosphorylation of ATM has been interpreted as an indication of ATM activation^{1,13}. Interference with autophosphorylation blocks ATM-mediated targeting of downstream substrates and the induction of cell cycle checkpoints. For example, inhibiting the activities of the MRE11–RAD50–NBS1 complex^{14–19}, 53BP1 (ref. 20), histone acetyltransferases^{21,22} and protein phosphatase-5 (ref. 23) all decrease the formation of S1981-phosphorylated ATM and concomitantly suppress other readouts of ATM kinase activity. Despite this correlation, a causative role for S1981 phosphorylation in the activation of ATM remains controversial. First, the ATM-S1981A mutant protein does not block dimer dissociation or ATM kinase activity in *in vitro* assays^{1,24}. Second, *Xenopus* ATM can be activated by DNA in the absence of autophosphorylation²⁵. Third, ATM is constitutively phosphorylated in certain cell lines without rendering ATM active towards downstream targets²⁶. Fourth, S1981 phosphorylation has no detectable effect on the oligomerization state of ATM: native ATM elutes as a single peak after gel filtration chromatography, and the position of this peak remains unchanged irrespective of irradiation in both human²⁷ and mouse (Supplementary Fig. 8) cells. Finally, pre-treatment of human cells with KU55933 (ref. 12), which blocks ATM autophosphorylation by competition with ATP, failed to inhibit the irradiation-induced increase in the kinase activity of immunoprecipitated ATM (Supplementary Fig. 9), further suggesting that autophosphorylation of ATM is not essential for its activation. Although we cannot exclude that the mechanism of ATM activation may operate differently in humans or that additional post-translational modifications may contribute to ATM activation^{21,23}, our findings demonstrate that phosphorylation of murine *Atm* at S1987 *in vivo* is not the signal that initially activates dormant *Atm* molecules, attracts *Atm* to sites of DNA damage, or facilitates its access to protein substrates.

Activation of ATM is dependent on MRE11–RAD50–NBS1-mediated DNA unwinding and recruitment of ATM to DNA ends^{2,3,10,24,28}. In addition, recent evidence suggests that ATM S1981 phosphorylation occurs in the vicinity of a DSB²⁹ subsequent to its initial recruitment³. We speculate that after *Atm* is converted to a catalytically active conformation at a DSB, a large pool of *Atm* becomes accessible to rapid phosphorylation as a result of the high local concentration of available S1987 targets in *Atm* proximal to the lesion. Nevertheless, this post-translational modification is not required for the activation or activity of murine *Atm* *in vivo*.

METHODS

Generation of mice. BAC RP24-122F10, which consists of a 160-kilobase (kb) insert, including 48.3 kb of sequence upstream and 17.9 kb of sequence downstream of the *Atm* initiation and stop codons, respectively, was used to generate transgenic mice. Targeting of the BAC was performed as described⁴. The presence of the transgene was determined by screening tail DNA using the following PCR primer pairs: mAtmF 5'-AGCACAACCACTGAATGC-3' and SP6R 5'-GTTTTCGCGATCTGCCGTTTC-3'; T7F 5'-TAATACGACTCACTATAGGG-3' and mAtmR 5'-CTCAGGTGGAATCTAACCTG-3', which amplify a product of 600 base pairs (bp) and 300 bp at the 5' and 3' ends of the insert, respectively. Transgenic founders were crossed to *Atm*^{+/-} mice.

Western blotting and immunofluorescence. For western blotting, primary antibodies were used at the following dilutions: anti-human ATM S1981p

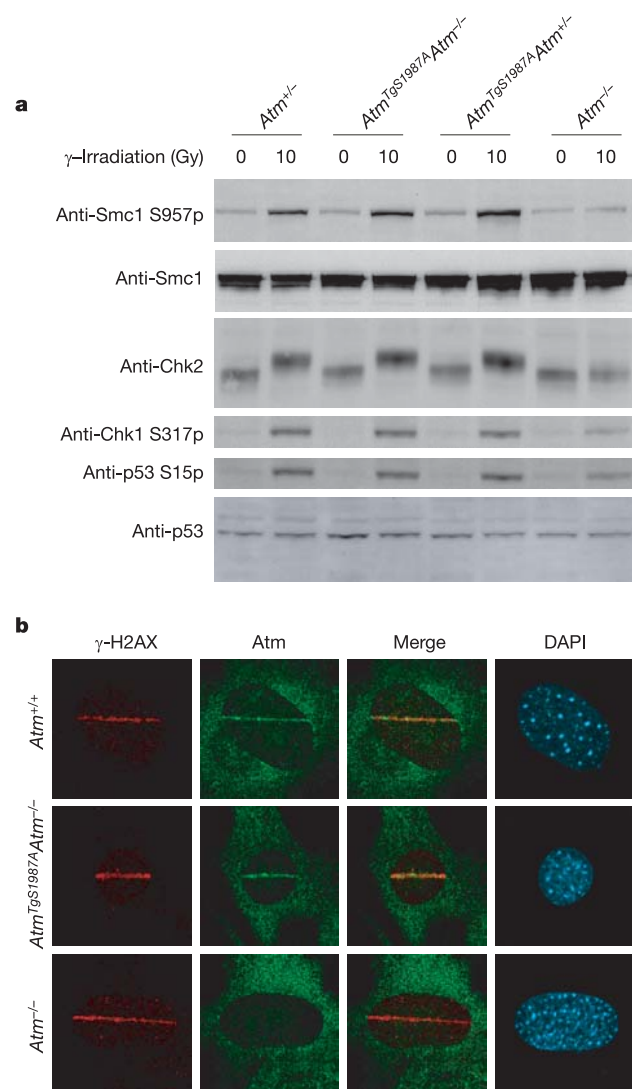


Figure 3 | S1987 phosphorylation is dispensable for *Atm* kinase activity and recruitment to DNA breaks. **a**, B cells were harvested for western blot analysis 30 min after no treatment (0 Gy) or after γ -irradiation (10 Gy). **b**, Distribution of *Atm* (green) in *Atm*^{+/-}, *Atm*^{-/-} and *Atm*^{TgS1987A}*Atm*^{+/-} fibroblasts 5 min after treatment with ultraviolet laser micro-irradiation. γ -H2AX (red) marks sites of DSBs.

(1:400, Rockland Immunochemicals), anti-human unphosphorylated ATM S1981 (1:2,500, Rockland Immunochemicals), anti-ATM 5C2 (1:500, Novus), anti-Smc1 S957p (1:500, Rockland Immunochemicals), anti-Smc1 (1:2,000, Novus), anti-p53 S15p (1:500, Cell Signaling), anti-p53 (1:1,000, Santa Cruz), anti-Chk1 317p (1:500, Bethyl Labs), anti-Chk2 (1:1,500, Upstate Biotechnology), anti-tubulin (1:10,000, Sigma) and anti-mouse Atm S1987p (1:500) was produced by immunization with the synthetic peptide SPTFEEGSpQGTTISS (Becton Dickinson). The association of Atm with chromatin after irradiation was determined using detergent extraction as described³⁰. Primary B cells and ear fibroblasts were irradiated with a 364-nm ultraviolet laser, and processed for immunofluorescence analysis as described²⁹.

Cell cycle checkpoints and chromosomal stability. B cells were challenged with different doses of irradiation after stimulation with LPS and IL-4, and intra-S-phase and G2/M checkpoints were measured as described¹⁹. Mitotic chromosome spreads were prepared after B-cell stimulation for 72 h, and metaphases with abnormalities were quantified⁶.

Additional materials and methods, including cell derivation, T-cell development, class switch recombination, immune complex assay, and histology are described in Supplementary Methods.

Received 5 April; accepted 25 July 2006.

Published online 13 August 2006.

- Bakkenist, C. J. & Kastan, M. B. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499–506 (2003).
- Falck, J., Coates, J. & Jackson, S. P. Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature* **434**, 605–611 (2005).
- You, Z., Chahwan, C., Bailis, J., Hunter, T. & Russell, P. ATM activation and its recruitment to damaged DNA require binding to the C terminus of Nbs1. *Mol. Cell. Biol.* **25**, 5363–5379 (2005).
- Yang, Y. & Sharan, S. K. A simple two-step, 'hit and fix' method to generate subtle mutations in BACs using short denatured PCR fragments. *Nucleic Acids Res.* **31**, e80 (2003).
- Bartkova, J. *et al.* ATM activation in normal human tissues and testicular cancer. *Cell Cycle* **4**, 838–845 (2005).
- Ramiro, A. R. *et al.* Role of genomic instability and p53 in AID-induced *c-myc-Igh* translocations. *Nature* **440**, 105–109 (2006).
- Liu, Q. *et al.* Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. *Genes Dev.* **14**, 1448–1459 (2000).
- Zachos, G., Rainey, M. D. & Gillespie, D. A. Chk1-deficient tumour cells are viable but exhibit multiple checkpoint and survival defects. *EMBO J.* **22**, 713–723 (2003).
- Takai, H. *et al.* Aberrant cell cycle checkpoint function and early embryonic death in Chk1^{-/-} mice. *Genes Dev.* **14**, 1439–1447 (2000).
- Kitagawa, R., Bakkenist, C. J., McKinnon, P. J. & Kastan, M. B. Phosphorylation of SMC1 is a critical downstream event in the ATM-NBS1-BRCA1 pathway. *Genes Dev.* **18**, 1423–1438 (2004).
- Bartek, J., Lukas, C. & Lukas, J. Checking on DNA damage in S phase. *Nature Rev. Mol. Cell Biol.* **5**, 792–804 (2004).
- Hickson, I. *et al.* Identification and characterization of a novel and specific inhibitor of the ataxia-telangiectasia mutated kinase ATM. *Cancer Res.* **64**, 9152–9159 (2004).
- Kozlov, S., Gueven, N., Keating, K., Ramsay, J. & Lavin, M. F. ATP activates ataxia-telangiectasia mutated (ATM) *in vitro*. Importance of autophosphorylation. *J. Biol. Chem.* **278**, 9309–9317 (2003).
- Uziel, T. *et al.* Requirement of the MRN complex for ATM activation by DNA damage. *EMBO J.* **22**, 5612–5621 (2003).
- Carson, C. T. *et al.* The Mre11 complex is required for ATM activation and the G2/M checkpoint. *EMBO J.* **22**, 6610–6620 (2003).
- Horejsi, Z. *et al.* Distinct functional domains of Nbs1 modulate the timing and magnitude of ATM activation after low doses of ionizing radiation. *Oncogene* **23**, 3122–3127 (2004).
- Cerosaletti, K. & Concannon, P. Independent roles for nibrin and Mre11-Rad50 in the activation and function of Atm. *J. Biol. Chem.* **279**, 38813–38819 (2004).
- Costanzo, V., Paull, T., Gottesman, M. & Gautier, J. Mre11 assembles linear DNA fragments into DNA damage signaling complexes. *PLoS Biol.* **2**, E110 (2004).
- Difilippantonio, S. *et al.* Role of Nbs1 in the activation of the Atm kinase revealed in humanized mouse models. *Nature Cell Biol.* **7**, 675–685 (2005).
- Mochan, T. A., Venere, M., DiTullio, R. A. Jr. & Halazonetis, T. D. 53BP1 and NFB1/MDC1-Nbs1 function in parallel interacting pathways activating ataxia-telangiectasia mutated (ATM) in response to DNA damage. *Cancer Res.* **63**, 8586–8591 (2003).
- Sun, Y., Jiang, X., Chen, S., Fernandes, N. & Price, B. D. A role for the Tip60 histone acetyltransferase in the acetylation and activation of ATM. *Proc. Natl Acad. Sci. USA* **102**, 13182–13187 (2005).
- Gupta, A. *et al.* Involvement of human MOF in ATM function. *Mol. Cell. Biol.* **25**, 5292–5305 (2005).
- Ali, A. *et al.* Requirement of protein phosphatase 5 in DNA-damage-induced ATM activation. *Genes Dev.* **18**, 249–254 (2004).
- Lee, J. H. & Paull, T. T. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science* **308**, 551–554 (2005).
- Dupre, A., Boyer-Chatenet, L. & Gautier, J. Two-step activation of ATM by DNA and the Mre11-Rad50-Nbs1 complex. *Nature Struct. Mol. Biol.* **13**, 451–457 (2006).
- Goldstine, J. V. *et al.* Constitutive phosphorylation of ATM in lymphoblastoid cell lines from patients with ICF syndrome without downstream kinase activity. *DNA Repair (Amst.)* **5**, 432–443 (2006).
- Goodarzi, A. A. & Lees-Miller, S. P. Biochemical characterization of the ataxia-telangiectasia mutated (ATM) protein from human cells. *DNA Repair (Amst.)* **3**, 753–767 (2004).
- Cerosaletti, K., Wright, J. & Concannon, P. Active role for nibrin in the kinetics of atm activation. *Mol. Cell. Biol.* **26**, 1691–1699 (2006).
- Kruhlak, M. J. *et al.* Changes in chromatin structure and mobility in living cells at sites of DNA double-strand breaks. *J. Cell Biol.* **172**, 823–834 (2006).
- Andegeko, Y. *et al.* Nuclear retention of ATM at sites of DNA double strand breaks. *J. Biol. Chem.* **276**, 38224–38230 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Sharan for assistance with BAC recombineering; S. Jay, D. Winkler, F. Van Laethem, M. Kruhlak and M. Eckhause for technical assistance; G. Smith for providing ATM and DNA-PKcs small molecule inhibitors; T. Paull for providing human monomeric ATM; and O. Fernandez-Capetillo and A. Singer for helpful suggestions on the manuscript. This work was supported by the Intramural research Program of the NIH, National Cancer Institute, Center for Cancer Research, National Institute of Aging, and the AT Children's Project (grant to A.N.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.N. (andre_nussenzweig@nih.gov).

LETTERS

Molecular identification of the CRAC channel by altered ion selectivity in a mutant of Orai

Andriy V. Yeromin^{1*}, Shenyuan L. Zhang^{1*}, Weihua Jiang¹, Ying Yu¹, Olga Safrina¹ & Michael D. Cahalan¹

Recent RNA interference screens have identified several proteins that are essential for store-operated Ca^{2+} influx and Ca^{2+} release-activated Ca^{2+} (CRAC) channel activity in *Drosophila* and in mammals, including the transmembrane proteins Stim (stromal interaction molecule)^{1,2} and Orai^{3–5}. Stim probably functions as a sensor of luminal Ca^{2+} content and triggers activation of CRAC channels in the surface membrane after Ca^{2+} store depletion^{1,6}. Among three human homologues of Orai (also known as *olf186-F*), *ORAI1* on chromosome 12 was found to be mutated in patients with severe combined immunodeficiency disease, and expression of wild-type Orai1 restored Ca^{2+} influx and CRAC channel activity in patient T cells³. The overexpression of Stim and Orai together markedly increases CRAC current^{5,7–9}. However, it is not yet clear whether Stim or Orai actually forms the CRAC channel, or whether their expression simply limits CRAC channel activity mediated by a different channel-forming subunit. Here we show that interaction between wild-type Stim and Orai, assessed by co-immunoprecipitation, is greatly enhanced after treatment with thapsigargin to induce Ca^{2+} store depletion. By site-directed mutagenesis, we show that a point mutation from glutamate to aspartate at position 180 in the conserved S1–S2 loop of Orai transforms the ion selectivity properties of CRAC current from being Ca^{2+} -selective with inward rectification to being selective for monovalent cations and outwardly rectifying. A charge-neutralizing mutation at the same position (glutamate to alanine) acts as a dominant-negative non-conducting subunit. Other charge-neutralizing mutants in the same loop express large inwardly rectifying CRAC current, and two of these exhibit reduced sensitivity to the channel blocker Gd^{3+} . These results indicate that Orai itself forms the Ca^{2+} -selectivity filter of the CRAC channel.

Orai and Orai1 possess four hydrophobic stretches that are predicted to span the membrane. On the basis of strategies used previously for several different ion channels, we made point mutations to investigate the most conserved loop between putative transmembrane segments (Supplementary Fig. 1a) and examined properties of ion selectivity, current–voltage rectification and block that are intimately associated with a pore-forming subunit. Wild-type or mutant Orai proteins were overexpressed together with wild-type Stim in S2 cells; messenger RNA and protein expression were verified by RT–PCR and western blotting (Fig. 1a and Supplementary Fig. 1b). In addition, by co-immunoprecipitation of the epitope-tagged wild-type proteins, we evaluated whether Stim and Orai are associated with each other before and after Ca^{2+} store depletion. When cells were cultured without stimulation, only limited interaction was seen. However, Ca^{2+} store depletion triggered by thapsigargin, which inhibits the SERCA (sarco/endoplasmic reticulum Ca^{2+} -ATPase) re-uptake pump, markedly increased Stim–Orai protein interaction when either protein was used to pull

down the other by co-immunoprecipitation (Fig. 1a). This result shows that both transfected proteins are expressed and, moreover, that they interact, either directly or as part of a complex, after Ca^{2+} store depletion to initiate CRAC channel activation. During whole-cell recording on the co-transfected cells, dialysis of a Ca^{2+} chelator (BAPTA) into the cytoplasm activated a very large Ca^{2+} -selective inward current, as shown previously⁵. Using a pipette solution that maintains normal Ca^{2+} store content to prevent spontaneous channel opening, we confirmed that addition of inositol-1,4,5-trisphosphate (IP_3) inside or thapsigargin (TG) outside resulted in greatly augmented CRAC current (Fig. 1b).

Expression of an Orai mutant with a glutamate-to-aspartate mutation at position 180 (Orai(E180D)), again together with wild-type Stim, generated a current that developed with the normal kinetics of CRAC current during passive Ca^{2+} store depletion, but also exhibited marked alterations in ion selectivity and rectification. In contrast to the normal inward CRAC current, Orai(E180D) expression produced a very large outward current and a smaller inward current that developed in parallel (Fig. 2a). Current–voltage (*I*–*V*) curves of the Orai(E180D)-induced current revealed an outwardly rectifying shape with a reversal potential of 6.0 ± 0.4 mV ($n = 14$ cells), in contrast to the characteristic inwardly rectifying

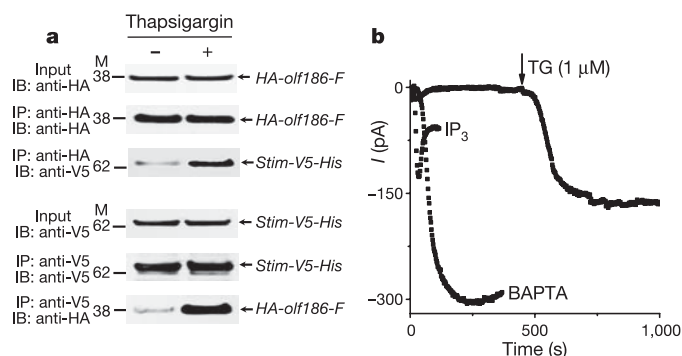


Figure 1 | Orai interacts with Stim and generates increased store-operated currents in S2 cells. **a**, Representative co-immunoprecipitation of Stim and Orai ($n = 3$). Input: total cell lysate, showing an equal amount of samples prepared for immunoprecipitation. Cells were transfected with HA-olf186-F (Orai) and Stim-V5-His (Stim) constructs, as indicated. M refers to molecular weight markers (kDa). **b**, Activation of CRAC current (at -130 mV) by three methods: IP_3 ($10 \mu\text{M}$ IP_3 added to high- Ca^{2+} pipette solution; current density -11.4 ± 3.1 pA pF $^{-1}$, $n = 5$); thapsigargin (high- Ca^{2+} pipette solution and thapsigargin added to bath; -7.6 ± 2.5 pA pF $^{-1}$, $n = 2$); or passive store-depletion (Ca^{2+} -free pipette solution; -28.7 ± 1.8 pA pF $^{-1}$, $n = 27$). All three methods produced substantially greater CRAC current density than passive store depletion in control cells (-2.5 ± 0.4 pA pF $^{-1}$, $n = 18$).

¹Department of Physiology & Biophysics and Center for Immunology, University of California, Irvine, California 92697, USA.

*These authors contributed equally to this work.

I - V curves with a reversal potential >50 mV of native or wild-type Orai-induced CRAC current (Fig. 2b).

In a series of ion-substitution experiments, we evaluated the ion selectivity properties of current induced by Orai(E180D). Substitution of Na^+ by choline while maintaining constant Ca^{2+} in the external solution—an experimental procedure that has little effect on native CRAC current¹⁰ or CRAC current with overexpressed wild-type Orai (Fig. 2c)—reduced the inward current and shifted the reversal potential to hyperpolarized potentials (Fig. 2d), indicating that the Orai(E180D)-induced inward current is carried predominantly by Na^+ rather than by Ca^{2+} . Next, different divalent cation species in the external solution were tested, resulting in a I - V pattern of wild-type Orai-induced CRAC current similar to native CRAC current described previously¹⁰. Ca^{2+} , Sr^{2+} and Ba^{2+} (20 mM in each case, while maintaining high external Na^+) resulted in increased inward current (Fig. 2e), indicating that each of these divalent cations is permeant, whereas Mg^{2+} is relatively impermeant. In contrast,

representative I - V curves of Orai(E180D)-induced current (shown in Fig. 2f with the same colour code for comparison with Fig. 2e) reveal an entirely different pattern. A tenfold increase of Ca^{2+} concentration ($[\text{Ca}^{2+}]$) in the external solution significantly decreased the outward current and did little to the small inward current. Substitution of 2 mM external Ca^{2+} by 20 mM Sr^{2+} had little effect; Mg^{2+} increased the current; and Ba^{2+} significantly increased the inward current while decreasing the outward current (Supplementary Fig. 2a). These results indicate that divalent cations block the outward current (carried by Cs^+) in the potency sequence $\text{Ca}^{2+} > \text{Ba}^{2+} > \text{Sr}^{2+} > \text{Mg}^{2+}$, and the inward current (carried by Na^+) in the potency sequence $\text{Ca}^{2+} \approx \text{Sr}^{2+} > \text{Mg}^{2+} > \text{Ba}^{2+}$. Together, these ion-substitution experiments indicate that the Orai(E180D) mutation transforms the CRAC channel from being selective for divalent cations to being a monovalent-selective channel that selects against, and indeed is blocked by, divalent cations.

CRAC channels share with voltage-gated Ca^{2+} channels the property of monovalent ion permeation when external divalent ions are removed^{11–15}. Upon withdrawal of external divalent cations, Na^+ carries significant current through CRAC channels, whereas Cs^+ is far less permeant^{10,14,16,17}. The monovalent CRAC current normally declines within 10 s, a process termed depotentiation¹⁵ that is due to removal of external Ca^{2+} . To check whether the Orai(E180D) mutation alters monovalent CRAC current, both wild-type- and Orai(E180D)-induced CRAC currents were recorded in divalent-free Na^+ and Cs^+ test solutions. Three clear differences can be discerned (Fig. 3). First, the inward Na^+ current was much larger in the Orai(E180D) mutant and did not depotentiate to the same extent as either wild-type Orai (Fig. 3a, b) or native CRAC current in S2 cells¹⁰. Second, in contrast to the relatively small Cs^+ current density in wild-type Orai (-1.6 ± 0.6 pA pF⁻¹, $n = 4$), Orai(E180D)-induced Cs^+ current density was much larger (-191 ± 46 pA pF⁻¹, $n = 4$), similar in amplitude to the Na^+ current (-238 ± 43 pA pF⁻¹, $n = 9$), and the large Cs^+ current did not depotentiate (Fig. 3b). Third, the measured reversal potential in Na^+ external solution was 7.9 ± 0.3 mV ($n = 9$) for Orai(E180D), whereas it was 52.4 ± 0.9 mV ($n = 5$) for wild-type Orai (Fig. 3c, d). Calculated from reversal potentials, the average permeability ratio $P_{\text{Na}}/P_{\text{Cs}}$ changed from 8.0 for wild-type Orai to 1.4 as a result of the E180D mutation, indicating an increase in pore diameter or altered electrostatic interactions with cations. I - V curves for Orai(E180D)-induced current changed shape with varying external $[\text{Ca}^{2+}]$ (Fig. 3e), revealing voltage-dependent block (Fig. 3f) that causes outward rectification in physiological saline. This examination of ion selectivity properties shows that a conservative mutation of glutamate to aspartate at position 180 of Orai transforms the CRAC channel from being Ca^{2+} -selective and inwardly rectifying to one that conducts Na^+ or Cs^+ and is outwardly rectifying due to voltage-dependent Ca^{2+} block.

Expression of the charge-neutralizing Orai(E180A) mutant completely abolished native CRAC current (Fig. 4a), indicating that alanine-containing subunits prevent ion conduction in hetero-multimeric CRAC channels by a dominant-negative action. In contrast, expression of Orai(D184A), Orai(D186A) or Orai(N188A) mutants with Stim resulted in increased CRAC current with unaltered I - V shape (Fig. 4b) and normal monovalent Na^+ current that exhibited depotentiation and low Cs^+ permeability (data not shown). Figure 4c summarizes the effect of each Orai mutant tested on CRAC current density. Expressed by itself or together with Stim, only Orai(E180A) potentially inhibited native CRAC current. The inward current density and I - V shape in cells expressing Orai(D184A), Orai(D186A) and Orai(N188A) mutants did not differ significantly from current observed in cells expressing wild-type Orai. Orai(E180D) was unique in producing very large outward currents as a result of altered ion selectivity.

Gd^{3+} very effectively suppresses native and wild-type Orai-induced CRAC current⁵. In addition to ion selectivity properties, block by Gd^{3+} was evaluated in all mutants (Fig. 4d). Two charge-neutralizing

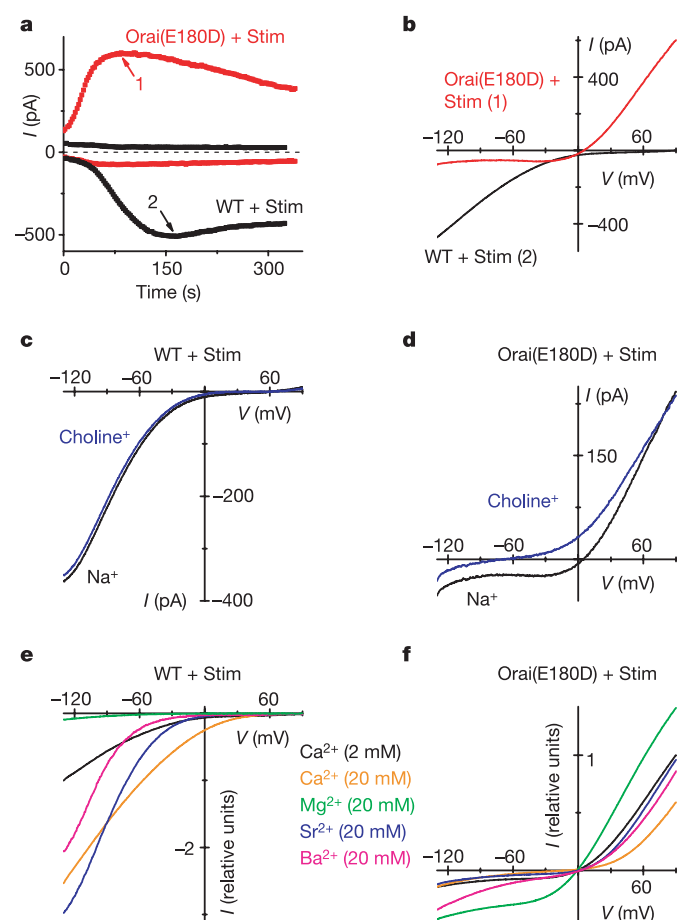


Figure 2 | Mutation E180D of Orai alters ion selectivity of the CRAC current. **a**, Time courses of inward current at -130 mV and outward current at 90 mV in representative cells overexpressing wild-type Orai (WT, black) or Orai(E180D) mutant (red). Ca^{2+} -free internal solution. **b**, I - V curves at times indicated in **a**. **c**, I - V curves of wild-type Orai-induced current in Ca^{2+} solution containing 160 mM Na^+ and 2 mM Ca^{2+} and in choline external solution with 1.1 mM Na^+ and 2 mM Ca^{2+} . (Complete solution recipes are indicated in Supplementary Table 1.) **d**, Representative I - V curves for Orai(E180D) with the same solutions as **c** ($n = 3$). **e**, Divalent cation selectivity of wild-type Orai CRAC current. I - V curves are normalized to current values at -130 mV in Ca^{2+} solution. Test solutions contained 20 mM test divalent and 124 mM Na^+ . External solutions for **e** and **f** are labelled according to colour. **f**, I - V curves (not leak-subtracted) for Orai(E180D) with the same divalent cations, normalized to currents at 90 mV in Ca^{2+} solution.

mutations, aspartate to alanine at positions 184 and 186, and, to a lesser degree, mutation of asparagine to alanine at position 188, significantly reduced the potency of Gd^{3+} block (Fig. 4d and Supplementary Fig. 2b). As shown previously⁵, wild-type Orai CRAC current was potentiated by 2-aminoethylidiphenyl borate (2-PB) at a low concentration (5 μM) and inhibited at a high (20 μM) concentration (Supplementary Fig. 3a), similar to effects of 2-APB on native CRAC current in Jurkat and in S2 cells^{10,18}. The effects of 2-APB on Orai(E180D)-induced CRAC current were complex but still showed potentiation and inhibition (Supplementary Fig. 3b). The time course of Orai(E180D) outward CRAC current showed an immediate decrease preceding potentiation by 5 μM 2-APB, and an immediate increase preceding inhibition during 20 μM application ($n = 3$). The pharmacological analysis provides support for the conclusion that the S1–S2 loop is involved in ion selectivity and block by Gd^{3+} , and may contribute to the complex effects of 2-APB on CRAC current.

Our results demonstrate that thapsigargin-triggered store

depletion dynamically strengthens an interaction between Stim and Orai, supporting a model for CRAC channel activation in which Stim serves as the Ca^{2+} sensor to detect store depletion and as the messenger to activate CRAC channels in the plasma membrane. More importantly, we conclude that Orai is a bona fide ion channel, based on the following: (1) RNA-interference-mediated knockdown of Orai expression suppresses thapsigargin-dependent Ca^{2+} influx and CRAC channel activity; (2) overexpression of Orai with or without Stim augments CRAC currents that exhibit biophysical properties identical to native CRAC current; and (3) mutations of negatively charged residues within the putative pore region of Orai significantly alter ion selectivity, current rectification and affinity to a charged channel blocker without altering channel activation kinetics. The marked alteration of these properties by a targeted point mutation provides definitive evidence that Orai embodies the pore-forming subunit of the CRAC channel.

The consensus sequence within the S1–S2 loop, 179VEVQLDxD186, contains the critical glutamate (bold) shown here to control ion selectivity properties of the CRAC channel, and two aspartates (underlined) that may help to attract Gd^{3+} (and Ca^{2+}) towards the pore. It is not similar to pore sequences found in other channels. Unlike the pore regions of voltage-gated Ca^{2+} (CaV) channels, which contain a relatively long loop and a ring of critical glutamates from different domains that form a high-affinity Ca^{2+} -binding site¹⁹, the

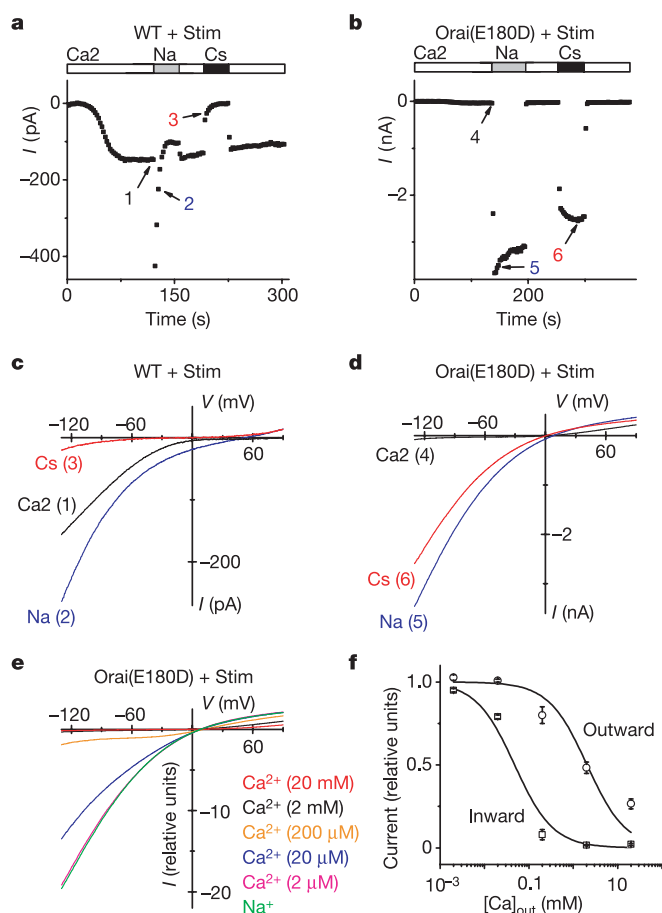


Figure 3 | Monovalent current in the absence of divalent ions exhibits altered ion selectivity in the Orai(E180D) mutant. **a**, Time course of inward current in a cell expressing wild-type Orai. Bars indicate external solution exchange. **b**, Same as **a** but for Orai(E180D). **c**, **d**, **e**, **f**, **I**–**V** curves for wild-type Orai at times indicated in **a**, **d**, **I**–**V** curves for Orai(E180D) at times indicated in **b**, **e**, **I**–**V** curves of Orai(E180D)-induced current in the presence of varying external $[\text{Ca}^{2+}]$. **f**, $[\text{Ca}^{2+}]$ dependence of Orai(E180D) CRAC current at -130 (squares) and 90 mV (circles), scaled to the current in divalent-free solution. Ca^{2+} -dependent block (mean $\pm$ s.e.m.) was fitted by the function $y = 1/(1 + [\text{Ca}^{2+}]/\text{IC}_{50})$, where IC_{50} is the calculated half-blocking Ca^{2+} concentration for inward current at -130 mV (48 ± 13 μM ; $n = 13, 8, 4, 6$ and 2 cells for 20, 2, 0.2, 0.02 and 0.002 mM free external $[\text{Ca}^{2+}]$, respectively) and outward current at 90 mV (2.05 ± 0.78 mM; $n = 13, 8, 4, 5$ and 3 cells for 20, 2, 0.2, 0.02 and 0.002 mM free external $[\text{Ca}^{2+}]$, respectively).

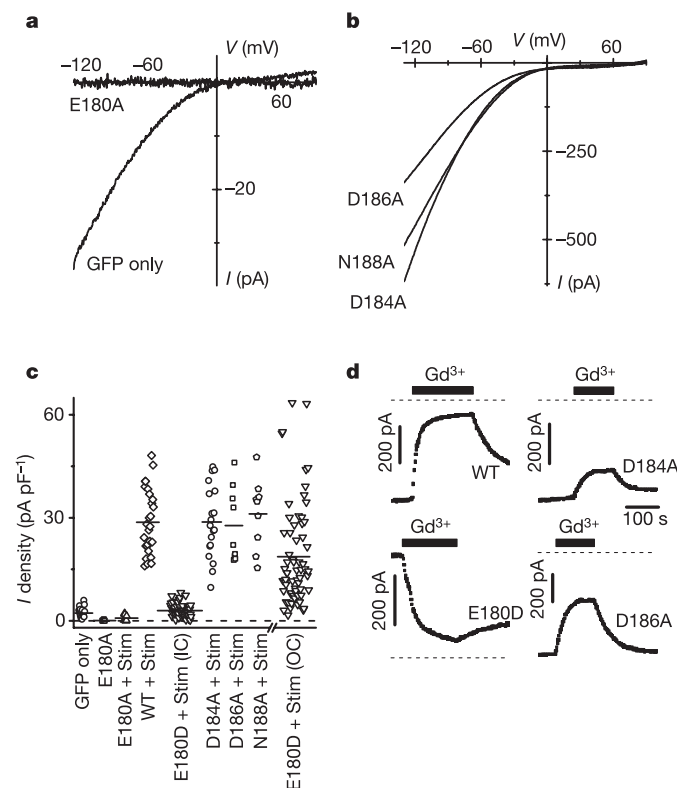


Figure 4 | Charge-neutralizing mutations in the S1–S2 loop. **a**, **I**–**V** curves in a cell expressing the Orai(E180A) mutant (without Stim) compared to control. **b**, **I**–**V** curves for Orai(D184A), Orai(D186A) and Orai(N188A) mutants. **c**, CRAC current density in transfected S2 cells. Each point represents the maximal CRAC current density (pA pF^{-1}) in a single cell, plotted as absolute values (from left to right): GFP-transfected control; Orai(E180A) ($P = 1.7 \times 10^{-4}$ relative to control); Orai(E180A) plus Stim ($P = 8.2 \times 10^{-3}$); wild-type Orai plus Stim; Orai(E180D) plus Stim, inward current; Orai(D184A) plus Stim; Orai(D186A) plus Stim; Orai(N188A) plus Stim; and Orai(E180D) plus Stim, outward current. **d**, Suppression of CRAC current by 5 nM Gd^{3+} in wild-type Orai and the D184A, E180D and D186A Orai mutants. Bars indicate time of Gd^{3+} application; dashed lines indicate the zero-current level.

putative pore sequence of Orai is very short, and the key residue for ion selectivity (E180) is adjacent to the putative S1 segment. Nevertheless, because withdrawal of external divalent ions reveals permeability to monovalent cations in both CaV^{11,12} and CRAC channels^{10,14,16,17}, it is possible that the CRAC channel ion-selectivity filter is also formed by a ring of glutamates and that the mechanism of Ca²⁺ permeation is similar, although the single-channel conductance and maximum permeant ion size of the CRAC channel selectivity filter are smaller than that of the CaV channel^{10,17,20,21}. Negatively charged side chains also contribute to Ca²⁺ selectivity of TRPV6; in this instance, aspartate (at position 541) is proposed to coordinate with Ca²⁺ ions and line the selectivity filter in a ring structure formed by four subunits²². The CRAC channel may be a multimer that includes several identical Orai subunits, as a non-conducting pore mutant (E180A) exerts a strong dominant-negative action on native CRAC current. Biochemical approaches and cysteine-scanning mutagenesis should be useful to elucidate better the unique pore architecture of the CRAC channel.

Note added in proof: While in proof, our attention was drawn to another paper reporting complementary findings regarding the control of ion selectivity in the CRAC channel²³. In our study, the critical glutamate residue is 180 in the Orai protein encoded by the Invitrogen S2 cell complementary DNA (GenBank accession number DQ503470, as described⁵). The corresponding residue is 178 in the *Drosophila* genome database (accession number AY071273), and 106 in the human Orai1 homologue (accession number BC015369).

METHODS

Molecular cloning and mutagenesis. Generation of pAc5.1/EGFP, pAc5.1/D-STIM, pAc5.1/D-STIM-V5-His and pAc5.1/olf186-F (GenBank accession number DQ503470) were as described⁵. Orai pore mutants were made by exchanging the corresponding codons (E180A, GAG to GCG; E180D, GAG to GAC; D184A, GAT to GCT; D186A, GAT to GCT; N188A, AAT to GCT; Supplementary Fig. 1) using the QuikChange site-directed mutagenesis kit (Stratagene). The pAc5.1/HA-olf186-F clone was made by adding the haemagglutinin tag via PCR and re-cloned into the *Xho*I and *Not*I sites of pAc5.1/V5-His B expression vector. Resulting clones were confirmed by sequencing. Description of the primers and conditions for cloning are available upon request. Both HA-olf186-F and Stim-V5-His were verified for normal function by whole-cell recording.

Whole-cell recording. Cells transfected (Amaxa) with Orai and Stim constructs were identified by fluorescence of co-transfected GFP. Patch-clamp experiments were performed at room temperature in the standard whole-cell recording, using conditions similar to those reported previously¹⁰. The membrane capacitance of S2 cells selected for recording was 11.6 ± 0.4 pF ($n = 131$ cells). Membrane potentials were corrected for a liquid junction potential of 10 mV between the pipette and bath solution. The series resistance (2–7 M Ω) was not compensated. The membrane potential was held at -10 mV, and 220-ms voltage ramps from -130 to 90 mV alternating with 220-ms pulses to -130 mV were delivered every 2 s. Only cells with high input resistance (>2 G Ω) were selected for recording. External and pipette solution recipes are listed in Supplementary Table 1. Analysed data are presented as mean $\pm$ s.e.m.

Co-immunoprecipitation. Two days after transfection, 5×10^6 cells were treated with either Ringer's solution or zero-Ca²⁺ Ringer's solution plus 1μ M thapsigargin using solutions described⁶ for 15 min at room temperature. Cells transfected with Stim-V5-His only or HA-olf186-F only were used as controls. Cell extracts were then prepared using RIPA lysis buffer (Upstate) according to the manufacturer's instructions. The extracts were pre-cleared with protein A/G beads (Pierce) and protein concentration was determined, using the Pierce BCA Protein Reagent Kit. The cell lysate was then diluted to approximately 1 mg ml^{-1} total protein with PBS and mixed with either anti-HA monoclonal antibodies (1μ g per 100μ g total protein, Santa Cruz) or anti-V5 monoclonal antibodies (1μ g per 100μ g total protein, Invitrogen) for 4 h at 4°C . Equal amounts of samples were mixed with either normal mouse IgG or vehicle as negative controls. Protein A/G beads were subsequently added (1μ l per 1μ g IgG) to rock overnight at 4°C . Proteins were eluted by boiling in 4 $\times$ sample buffer (Invitrogen). Samples were resolved by SDS-PAGE and analysed by standard western blotting techniques. Anti-HA monoclonal antibodies (Santa Cruz) were used at a dilution of 1:200. Anti-V5 monoclonal antibodies (Invitrogen) were used at a dilution of 1:2,500. Proteins were detected by developing with the SuperSignal (Pierce) detection system. Samples treated with normal mouse IgG

or vehicle showed no detectable signal for both Stim-V5-His and HA-olf186-F (data not shown), indicating that the immunoprecipitation and co-immunoprecipitation are specific.

Received 6 May; accepted 25 July 2006.

Published online 20 August 2006.

- Liou, J. *et al.* STIM is a Ca²⁺ sensor essential for Ca²⁺-store-depletion-triggered Ca²⁺ influx. *Curr. Biol.* **15**, 1235–1241 (2005).
- Roos, J. *et al.* STIM1, an essential and conserved component of store-operated Ca²⁺ channel function. *J. Cell Biol.* **169**, 435–445 (2005).
- Feske, S. *et al.* A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. *Nature* **441**, 179–185 (2006).
- Vig, M. *et al.* CRACM1 is a plasma membrane protein essential for store-operated Ca²⁺ entry. *Science* **312**, 1220–1223 (2006).
- Zhang, S. L. *et al.* Genome-wide RNAi screen of Ca²⁺ influx identifies genes that regulate Ca²⁺ release-activated Ca²⁺ channel activity. *Proc. Natl Acad. Sci. USA* **103**, 9357–9362 (2006).
- Zhang, S. L. *et al.* STIM1 is a Ca²⁺ sensor that activates CRAC channels and migrates from the Ca²⁺ store to the plasma membrane. *Nature* **437**, 902–905 (2005).
- Mercer, J. C. *et al.* Large store-operated calcium-selective currents due to co-expression of Orai1 or Orai2 with the intracellular calcium sensor, Stim1. *J. Biol. Chem.* doi:10.1074/jbc.M604589200 (published online 28 June 2006).
- Peinelt, C. *et al.* Amplification of CRAC current by STIM1 and CRACM1 (Orai1). *Nature Cell Biol.* **8**, 771–773 (2006).
- Soboloff, J. *et al.* Orai1 and STIM1 reconstitute store-operated calcium channel function. *J. Biol. Chem.* **281**, 20661–20665 (2006).
- Yeromin, A. V., Roos, J., Stauderman, K. A. & Cahalan, M. D. A store-operated calcium channel in *Drosophila* S2 cells. *J. Gen. Physiol.* **123**, 167–182 (2004).
- Almers, W. & McCleskey, E. W. Non-selective conductance in calcium channels of frog muscle: calcium selectivity in a single-file pore. *J. Physiol. (Lond.)* **353**, 585–608 (1984).
- Hess, P. & Tsien, R. W. Mechanism of ion permeation through calcium channels. *Nature* **309**, 453–456 (1984).
- Hoth, M. & Penner, R. Calcium release-activated calcium current in rat mast cells. *J. Physiol. (Lond.)* **465**, 359–386 (1993).
- Lepple-Wienhues, A. & Cahalan, M. D. Conductance and permeation of monovalent cations through depletion-activated Ca²⁺ channels (ICRAC) in Jurkat T cells. *Biophys. J.* **71**, 787–794 (1996).
- Zweifach, A. & Lewis, R. S. Calcium-dependent potentiation of store-operated calcium channels in T lymphocytes. *J. Gen. Physiol.* **107**, 597–610 (1996).
- Kozak, J. A., Kerschbaum, H. H. & Cahalan, M. D. Distinct properties of CRAC and MIC channels in RBL cells. *J. Gen. Physiol.* **120**, 221–235 (2002).
- Bakowski, D. & Parekh, A. B. Monovalent cation permeability and Ca²⁺ block of the store-operated Ca²⁺ current I_{CRAC} in rat basophilic leukemia cells. *Pflügers Arch.* **443**, 892–902 (2002).
- Prakriya, M. & Lewis, R. S. Potentiation and inhibition of Ca²⁺ release-activated Ca²⁺ channels by 2-aminoethyl-diphenyl borate (2-APB) occurs independently of IP₃ receptors. *J. Physiol. (Lond.)* **536**, 3–19 (2001).
- Ellinor, P. T., Yang, J., Sather, W. A., Zhang, J. F. & Tsien, R. W. Ca²⁺ channel selectivity at a single locus for high-affinity Ca²⁺ interactions. *Neuron* **15**, 1121–1132 (1995).
- Prakriya, M. & Lewis, R. S. Separation and characterization of currents through store-operated CRAC channels and Mg²⁺-inhibited cation (MIC) channels. *J. Gen. Physiol.* **119**, 487–507 (2002).
- Zweifach, A. & Lewis, R. S. Mitogen-regulated Ca²⁺ current of T lymphocytes is activated by depletion of intracellular Ca²⁺ stores. *Proc. Natl Acad. Sci. USA* **90**, 6295–6299 (1993).
- Voets, T., Janssens, A., Droogmans, G. & Nilius, B. Outer pore architecture of a Ca²⁺-selective TRP channel. *J. Biol. Chem.* **279**, 15223–15230 (2004).
- Prakriya, M. *et al.* Orai1 is an essential pore subunit of the CRAC channel. *Nature* advance online publication, doi:10.1038/nature05122 (20 August 2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Forrest for assistance in cell culture and G. Chandy for use of molecular reagents in his laboratory. This work was supported by a grant from the National Institutes of Health (M.D.C.), by a fellowship from the George E. Hewitt Foundation (S.L.Z.), and by a Scientist Development Grant from the American Heart Association (Y.Y.).

Author Contributions A.V.Y. was responsible for all patch-clamp experiments and analysis. S.L.Z. was responsible for all molecular biology and biochemistry experiments, with the assistance of W.J. Y.Y. and O.S. performed RT-PCR. M.D.C. provided advice and overall direction, and supervised project planning and execution.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.D.C. (mcahalan@uci.edu).

LETTERS

Orai1 is an essential pore subunit of the CRAC channel

Murali Prakriya^{1*}, Stefan Feske^{2*}, Yousang Gwack^{2*}, Sonal Srikanth², Anjana Rao² & Patrick G. Hogan²

Stimulation of immune cells causes depletion of Ca^{2+} from endoplasmic reticulum (ER) stores, thereby triggering sustained Ca^{2+} entry through store-operated Ca^{2+} release-activated Ca^{2+} (CRAC) channels, an essential signal for lymphocyte activation and proliferation^{1,2}. Recent evidence indicates that activation of CRAC current is initiated by STIM proteins, which sense ER Ca^{2+} levels through an EF-hand located in the ER lumen and relocate upon store depletion into puncta closely associated with the plasma membrane^{3–5}. We and others recently identified *Drosophila* Orai and human Orai1 (also called TMEM142A) as critical components of store-operated Ca^{2+} entry downstream of STIM^{6–8}. Combined overexpression of Orai and Stim in *Drosophila* cells⁸, or Orai1 and STIM1 in mammalian cells^{9–11}, leads to a marked increase in CRAC current. However, these experiments did not establish whether Orai is an essential intracellular link between STIM and the CRAC channel, an accessory protein in the plasma membrane, or an actual pore subunit. Here we show that Orai1 is a plasma membrane protein, and that CRAC channel function is sensitive to mutation of two conserved acidic residues in the transmembrane segments. E106D and E190Q substitutions in transmembrane helices 1 and 3, respectively, diminish Ca^{2+} influx, increase current carried by monovalent cations, and render the channel permeable to Cs^+ . These changes in ion selectivity provide strong evidence that Orai1 is a pore subunit of the CRAC channel.

To examine whether human Orai1 is expressed at the cell surface, HEK293 cell lines stably expressing Flag-tagged Orai1 were treated with sulphydryl-NHS-biotin, a cell-impermeant biotinylation reagent. Western analysis of anti-Flag immunoprecipitates with horseradish-peroxidase-conjugated streptavidin showed that Flag-tagged Orai1 was labelled with biotin (Fig. 1a). Parallel control experiments, in which cells were treated with tunicamycin to inhibit ER processing and cell surface expression of glycoproteins, showed that intracellular Orai1 is not labelled by this protocol. The surface localization of Orai1 was confirmed by immunocytochemical staining of haemagglutinin (HA)-tagged Orai1, with the epitope inserted between residues 214 and 217 in the predicted TM3–TM4 loop, in intact cells (Fig. 1c and Supplementary Fig. 1).

The TM3–TM4 extracellular loop is marked by the presence of a consensus N-linked glycosylation site, N223, that is glycosylated in fully processed Orai1 (Fig. 1b). Glycosylation at this site requires that N223 is in the ER lumen during glycoprotein maturation, and is extracellular after protein incorporation into the plasma membrane. In contrast to the HA epitope inserted into the TM3–TM4 loop, epitope tags located at the amino and carboxy termini of Orai1 were detected only after cell permeabilization (Fig. 1d and Supplementary Fig. 1)^{6,7}. These data are consistent with the prediction that Orai1 is a four-transmembrane-segment protein

with its N and C termini located in the cell cytoplasm (Fig. 1e). The extracellular location of the predicted short TM1–TM2 loop remains to be directly verified.

Conserved acidic residues have a key role in ion permeation through known Ca^{2+} channels^{12,13}. Orai-family proteins contain four glutamate residues that are conserved across species—E178, E221, E245 and E262 in *Drosophila* Orai—and a cluster of conserved acidic or amide side chains represented in *Drosophila* Orai by Q180, D182, D184 and N186 (Supplementary Fig. 2). We developed an assay to examine the function of expressed recombinant Orai proteins with essentially no contribution from endogenous Orai (Supplementary Methods and Supplementary Fig. 3). In this assay, wild-type Orai and the E221Q mutant (Orai(E221Q)) supported thapsigargin-induced Ca^{2+} influx comparable to that in control cells, and the Orai(E245Q) mutant supported a partial Ca^{2+} influx (Fig. 2a, b). In contrast, the Orai(E178Q) and Orai(E262Q) mutants did not support store-operated Ca^{2+} entry (Fig. 2b), even though the proteins were expressed at levels comparable to the wild-type and Orai(E221Q) mutant proteins, respectively (Supplementary Fig. 3).

We next examined the effect of amino acid substitutions at E106 and E190 of human Orai1, which correspond to E178 and E262 of *Drosophila* Orai, by expressing mutated proteins in CRAC-current- and Ca^{2+} -influx-deficient T cells and fibroblasts from patients with severe combined immunodeficiency (SCID)^{6,14}. Cells were stably transduced with complementary DNAs encoding E-to-A, E-to-D and E-to-Q mutants in a bicistronic IRES-GFP vector, and store-operated Ca^{2+} influx was evaluated in GFP⁺ cells (Fig. 2c, d and Supplementary Figs 4 and 5). The Orai1(E106A) and Orai1(E106Q) mutants failed to restore significant Ca^{2+} influx, whereas the Orai1(E106D) mutant restored Ca^{2+} influx partially. The Orai1(E190Q) mutant was marginally effective in T cells, whereas the Orai1(E190A) and Orai1(E190D) mutants fully restored Ca^{2+} influx. In all cases the residual Ca^{2+} influx resembled CRAC influx in that it was dependent on store depletion and sensitive to inhibition by a submicromolar concentration of La^{3+} (Supplementary Fig. 6). All mutants were expressed normally in the plasma membrane of HEK293 cells and fibroblasts (Supplementary Fig. 5 and data not shown).

To define the effects of Orai1 mutations more precisely, we examined the CRAC current in SCID T cells expressing mutant Orai1 proteins using patch-clamp recording. Average current at -100 mV was smaller in T cells expressing the Orai1(E106D) mutant (0.46 ± 0.06 pA pF⁻¹ in 20 mM extracellular Ca^{2+} concentration [$\text{Ca}^{2+}]_o$; $n = 10$ cells) (Fig. 3a) than in cells expressing wild-type Orai1 (2.18 ± 0.22 pA pF⁻¹; $n = 9$ cells)⁶, but significantly larger than the currents in untransfected SCID T cells or in SCID T cells expressing the non-functional mutant Orai1(R91W) (Supplementary

¹Department of Molecular Pharmacology and Biological Chemistry, Northwestern University, Feinberg School of Medicine, Chicago, Illinois 60611, USA. ²Harvard Medical School and the CBR Institute for Biomedical Research, 200 Longwood Avenue, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

Fig. 7). The current–voltage (I – V) relationship in 20 mM $[Ca^{2+}]_o$ exhibited a reversal potential of 36 ± 5 mV ($n = 11$ cells). Lowering $[Ca^{2+}]_o$ to 2 mM shifted the reversal potential to 13 ± 5 mV ($n = 7$ cells) (Fig. 3a). In contrast, cells expressing the Orai1(E190D) mutant, which showed apparently normal Ca^{2+} influx (Fig. 2d), had an I – V relationship that was inwardly rectifying with a reversal potential more positive than +80 mV (Fig. 3b), similar to cells expressing native CRAC channels or wild-type Orai1 (Fig. 3d)^{6,15,16}.

The reversal potentials observed in cells expressing the

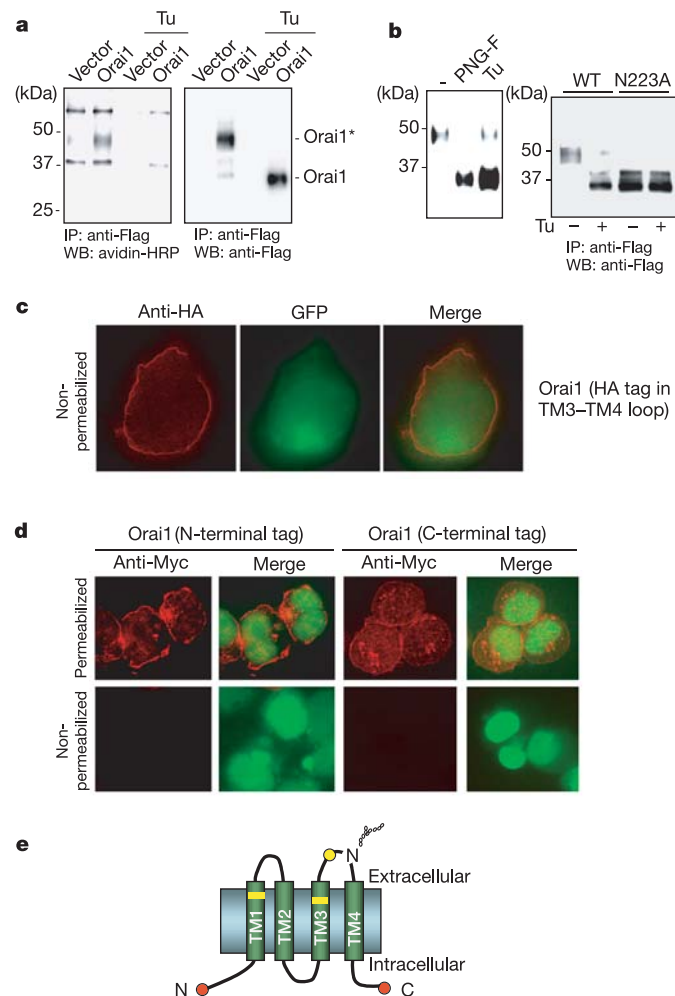


Figure 1 | Orai1 is a four-transmembrane protein expressed at the cell surface and with its N and C termini located in the cytoplasm. **a**, HEK293 cells expressing Flag–Orai1, left untreated or treated with tunicamycin (Tu), were surface-labelled with sulfo-NHS-biotin. Anti-Flag immunoprecipitates were probed for biotinylated Flag–Orai1 (left) and for total Orai1 (right). Orai1*, glycosylated Orai1. **b**, Anti-Flag immunoprecipitates from HEK293 cells expressing Flag–Orai1 were digested with peptide N-glycosidase F (PNG-F) and compared with samples from tunicamycin-treated cells (left). HEK293 cells expressing Flag-tagged wild-type Orai1 or Orai1(N223A) were treated with tunicamycin overnight (+) or left untreated (–) (right). **c**, Intact HEK293 cells expressing Orai1 with an HA tag in the predicted second extracellular loop were stained with anti-HA antibodies. **d**, SCID T cells expressing N- or C-terminally Myc-tagged Orai1 were permeabilized or left unpermeabilized, then stained with anti-Myc antibodies. **e**, Schematic representation of the membrane topology of Orai1 inferred from the hydropathy analysis of Orai1 and the above experiments. The sites of N-glycosylation (branched chain) (**b**), the HA tag (yellow circle) (**c**), and the Myc-epitope tags (red circles) (**d**) are indicated. The positions of residues E106 and E190 (corresponding to E178 and E262 in *Drosophila* Orai) in the transmembrane segments are highlighted in yellow.

Orai1(E106D) mutant were significantly less positive than those of native CRAC channels^{15,16}. Furthermore, significant outward currents were observed at potentials above the reversal potential (Fig. 3a), unlike what is seen for native CRAC current^{15,16}. These changes imply that the CRAC channel has lost some of its very high selectivity for Ca^{2+} over monovalent cations. In support of this interpretation, replacement of an impermeant ion, NMDG⁺, with Na^+ increased the current, indicating that in 20 mM Ca^{2+} Ringer's solution, Na^+ carried a significant amount of the inward current (Fig. 3e). This is a striking difference from the Orai1(E190D) mutant (Fig. 3f) and from native CRAC channels, which conduct very little Na^+ in the presence of millimolar concentrations of extracellular Ca^{2+} due to Ca^{2+} -mediated block of the Na^+ influx^{15,16}.

Like voltage-gated Ca^{2+} channels¹², CRAC channels achieve high Ca^{2+} selectivity because Ca^{2+} bound within the pore blocks the permeation of Na^+ and other monovalent cations through electrostatic repulsion. In the absence of external divalent ions, CRAC channels become freely permeable to monovalent ions^{16–19}, and these currents are blocked by Ca^{2+} with half-maximal block at 4–20 μ M Ca^{2+} (refs 17, 18, 20, 21). In contrast, in cells expressing the Orai1(E106D) protein, no detectable inhibition of the Na^+ current was observed with Ca^{2+} concentrations of 20–60 μ M, and half-maximal block was achieved only with 220 μ M Ca^{2+} (Fig. 3h, j, Supplementary Fig. 8 and data not shown). Thus, the E106D substitution results in a tenfold lower sensitivity of the CRAC channel Na^+ current to blockade by Ca^{2+} . These data argue that E106 coordinates Ca^{2+} at a high-affinity binding site within the pore of the CRAC channel.

A distinctive characteristic of native CRAC channels is their low permeability to Cs^+ , which confers a reversal potential of around +50 mV when internal Cs^+ and external Na^+ are the sole current carriers^{17–19,22}. In contrast, in cells expressing the Orai1(E106D)

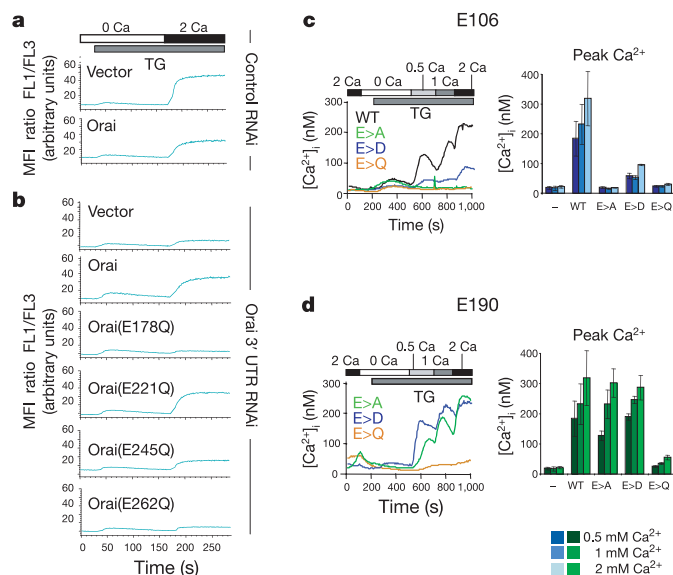


Figure 2 | The conserved glutamate residues in transmembrane segments TM1 and TM3 have an important role in store-operated Ca^{2+} influx.

a, **b**, Store-operated Ca^{2+} influx in S2R⁺ cells expressing *Drosophila* Orai or its mutants after treatment with control RNAi (**a**) or after depletion of endogenous Orai by RNAi (**b**). Cells were treated with thapsigargin (TG) in the absence of extracellular Ca^{2+} followed by re-addition of 2 mM Ca^{2+} . Changes in intracellular Ca^{2+} were monitored as the mean fluorescence intensity (MFI) ratio of the flow cytometric signals FL1 (Fluo-4) and FL3 (Fura Red). **c**, **d**, Ca^{2+} influx and averaged peak Ca^{2+} responses in SCID T cells expressing empty vector (–), wild-type Orai1, or Orai1 mutants, after store depletion with thapsigargin and re-addition of 0.5 mM, 1 mM and 2 mM Ca^{2+} . Histograms show the averaged peak Ca^{2+} responses ($\pm$ s.e.m., where appropriate) from two to five experiments per mutant.

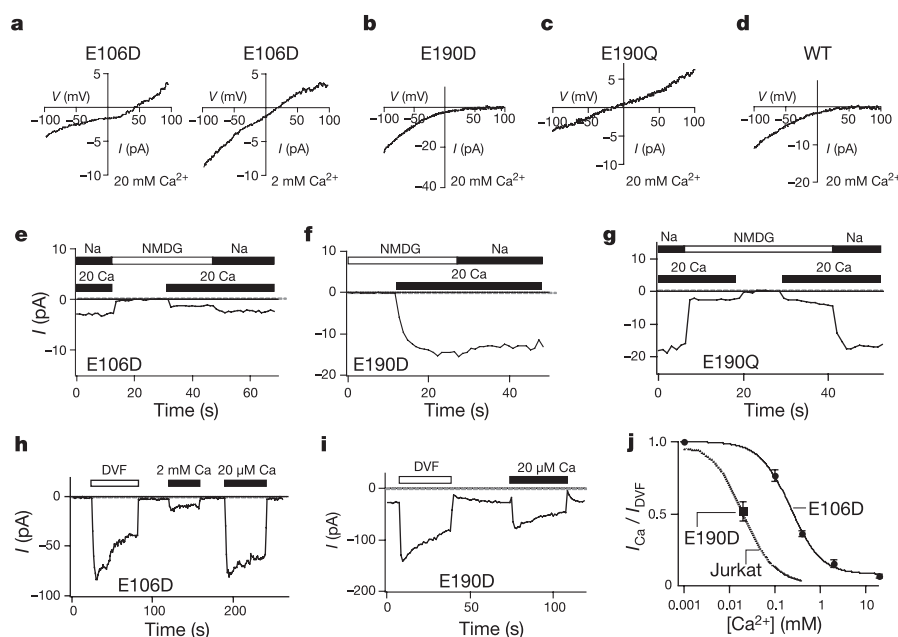


Figure 3 | E106D and E190Q mutations in the transmembrane segments of human Orai1 alter the ion selectivity of CRAC channels. **a–d**, I – V relations recorded in standard Ringer's solutions from SCID T cells expressing mutated or wild-type Orai1. **e–g**, Effect on CRAC current of changing to an extracellular solution with NMDG⁺, an impermeant cation, as the sole extracellular cation, then adding back Ca²⁺ (20 mM), then replacing NMDG⁺ by Na⁺. In these panels, the zero-current level (dashed line) is defined as the current in NMDG⁺ solution lacking Na⁺ and Ca²⁺. **h, i**, Effect of 20 μ M Ca²⁺ on Na⁺ current through CRAC channels in cells expressing

Orai1 (E106D) or Orai1 (E190D). Standard Ringer's solution containing 20 mM Ca²⁺ was replaced by solutions with [Ca²⁺]_o as indicated. DVF, divalent ion-free solution. **j**, Concentration dependence of Na⁺-CRAC current block by extracellular Ca²⁺ in cells expressing Orai1 (E106D). The solid line is a Hill equation fit to the data with $K_i = 0.22$ mM and Hill coefficient of 1.2. The dotted curve shows the corresponding inhibition curve for Jurkat T cells²¹. A single point representing block at 20 μ M Ca²⁺ (residual current $52 \pm 7\%$; $n = 4$ cells) is plotted for Orai1 (E190D). Values are mean $\pm$ s.e.m.

mutant, the current under these conditions reversed at ~ 0 mV (2 ± 2 mV, $n = 6$ cells), indicating approximately equal permeabilities to internal Cs⁺ and external Na⁺ (Fig. 4a). Extracellular application of Cs⁺ in a divalent cation-free solution elicited a large inward current in cells expressing Orai1 (E106D), providing direct evidence that the CRAC channel in these cells is permeable to Cs⁺ (Fig. 4d). The increase in Cs⁺ currents could, in principle, be due to changes in

electrostatic or chemical interactions of Cs⁺ with sites in the channel, but, given the exceptionally narrow pore of the CRAC channel^{17,21}, a subtle widening of the narrowest region of the pore could also contribute. The Orai1 (E190D) mutation did not alter the low permeability of CRAC channels to Cs⁺ (Fig. 4b, e).

CRAC current reconstituted with the Orai1 (E106D) mutant protein thus differs in at least three ways from CRAC current in T cells and rat basophilic leukaemia (RBL) cells and from CRAC current reconstituted with wild-type Orai1 (refs 6, 15, 16). First, there is reduced blockade of monovalent ion currents by Ca²⁺, so that ~ 220 μ M Ca²⁺ rather than ~ 4 – 20 μ M Ca²⁺ (refs 17, 18, 21) is required to achieve half-maximal block. Second, unlike native CRAC channels^{15,16}, the altered channel does not discriminate effectively between Na⁺ and Ca²⁺, and carries substantial Na⁺ current in physiological solutions. Third, Cs⁺ shows an increased ability to pass through the channel, with estimated permeability ratio $P_{Cs}/P_{Na} \approx 1$. In contrast, estimates of P_{Cs}/P_{Na} for native CRAC channels have ranged from 0.08 to 0.14 (refs 17, 19, 22).

Cells expressing Orai1 (E190Q) showed an even more striking loss of Ca²⁺ selectivity, as judged from the shift of reversal potential to -13 ± 1.5 mV ($n = 4$ cells) in the standard extracellular solution (Fig. 3c), and substantial inward Na⁺ current in the presence of 20 mM Ca²⁺ (Fig. 3g). With internal Cs⁺ and external Na⁺ as the sole current carriers, the CRAC current in these cells reversed at -6 ± 0.5 mV ($n = 4$ cells) (Fig. 4c), indicating a high relative permeability to Cs⁺. These data indicate that E190 also has a role in Ca²⁺ permeation through the CRAC channel.

The altered channel properties observed with E106D and E190Q substitutions in Orai1 parallel those produced by replacing key glutamate residues in voltage-dependent Ca²⁺ channels and in the relatively Ca²⁺-selective TRP channels TRPV5 and TRPV6 (refs 12, 13, 23–26). Because Orai1 is not obviously related to these channels, the similarities presumably reflect common physical principles that confer selective permeability to Ca²⁺ (refs 27, 28).

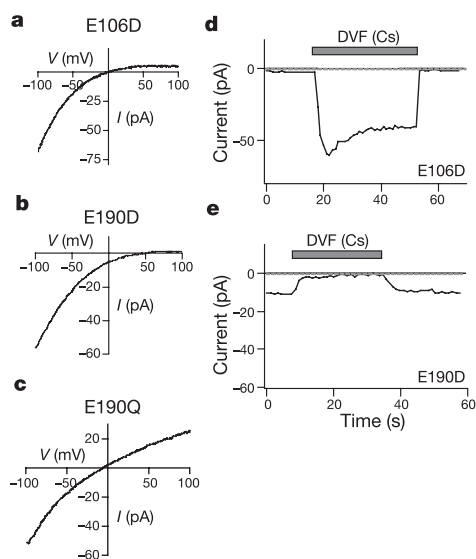


Figure 4 | E106D and E190Q mutations of human Orai1 alter the Cs⁺ selectivity of CRAC channels. **a–c**, I – V relations recorded in DVF solution (Na⁺ outside, Cs⁺ inside) in SCID T cells expressing mutated Orai1 proteins. **d, e**, Effect of replacing standard Ringer's solution with a Cs⁺-based DVF solution.

Human Orai1 is a plasma membrane protein, probably comprising four transmembrane helices and possessing highly conserved glutamate residues in transmembrane helices 1 and 3. Calcium influx through the CRAC channel reconstituted with Orai1 is extremely sensitive to replacement of E106 in transmembrane helix 1. The conservative substitution E106D weakens the interaction of Ca^{2+} with the channel and alters monovalent and divalent ion permeation. Qualitatively similar effects are produced by the substitution E190Q. These findings indicate that Orai1 is a pore subunit of the CRAC channel.

Note added in proof: *Drosophila* Orai protein residues are numbered according to the protein RE30427, encoded by the complementary DNA we used in our experiments (GenBank accession number AY071273). Recently, we became aware of a related paper³⁰ that reports on mutations in Orai encoded by a different cDNA (accession number DQ503470). Residue E180 in that paper corresponds to residue E178 in our numbering, and to residue E106 in human Orai1 (accession number BC015369). The two papers are in agreement on the effects of amino-acid substitutions at this position.

METHODS

See Supplementary Information for detailed methods.

Wild-type or mutant Orai1 proteins were stably expressed in human T-cell and fibroblast lines using a bicistronic expression vector that also expressed GFP. Single-cell Ca^{2+} imaging of cells expressing GFP was performed as described²⁹. Whole-cell recording of CRAC currents followed standard methods¹⁹.

Drosophila cell lines stably expressing wild-type or mutant Orai were generated by co-transfecting S2R⁺ cells with an Orai expression vector and a hygromycin resistance gene, followed by selection with hygromycin. Ca^{2+} influx in S2 R⁺ cells was measured by flow cytometry⁶.

Received 26 June; accepted 2 August 2006.

Published online 20 August 2006.

- Parekh, A. B. & Putney, J. W. Jr. Store-operated calcium channels. *Physiol. Rev.* **85**, 757–810 (2005).
- Lewis, R. S. Calcium signalling mechanisms in T lymphocytes. *Annu. Rev. Immunol.* **19**, 497–521 (2001).
- Roos, J. et al. STIM1, an essential and conserved component of store-operated Ca^{2+} channel function. *J. Cell Biol.* **169**, 435–445 (2005).
- Zhang, S. L. et al. STIM1 is a Ca^{2+} sensor that activates CRAC channels and migrates from the Ca^{2+} store to the plasma membrane. *Nature* **437**, 902–905 (2005).
- Liou, J. et al. STIM is a Ca^{2+} sensor essential for Ca^{2+} -store-depletion-triggered Ca^{2+} influx. *Curr. Biol.* **15**, 1235–1241 (2005).
- Feske, S. et al. A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. *Nature* **441**, 179–185 (2006).
- Vig, M. et al. CRACM1 is a plasma membrane protein essential for store-operated Ca^{2+} entry. *Science* **312**, 1220–1223 (2006).
- Zhang, S. L. et al. Genome-wide RNAi screen of Ca^{2+} influx identifies genes that regulate Ca^{2+} release-activated Ca^{2+} channel activity. *Proc. Natl Acad. Sci. USA* **103**, 9357–9362 (2006).
- Peinelt, C. et al. Amplification of CRAC current by STIM1 and CRACM1 (Orai1). *Nature Cell Biol.* **8**, 771–773 (2006).
- Soboloff, J. et al. Orai1 and STIM reconstitute store-operated calcium channel function. *J. Biol. Chem.* **281**, 20661–20665 (2006).
- Mercer, J. C. et al. Large store-operated calcium-selective currents due to co-expression of Orai1 or Orai2 with the intracellular calcium sensor, Stim1. *J. Biol. Chem.* doi:10.1074/jbc.M604589200 (published online 28 June 2006).
- Sather, W. A. & McCleskey, E. W. Permeation and selectivity in calcium channels. *Annu. Rev. Physiol.* **65**, 133–159 (2003).
- Owsianik, G., Talavera, K., Voets, T. & Nilius, B. Permeation and selectivity of TRP channels. *Annu. Rev. Physiol.* **68**, 685–717 (2006).
- Feske, S., Prakriya, M., Rao, A. & Lewis, R. S. A severe defect in CRAC Ca^{2+} channel activation and altered K^{+} channel gating in T cells from immunodeficient patients. *J. Exp. Med.* **202**, 651–662 (2005).
- Zweifach, A. & Lewis, R. S. Mitogen-regulated Ca^{2+} current of T lymphocytes is activated by depletion of intracellular Ca^{2+} stores. *Proc. Natl Acad. Sci. USA* **90**, 6295–6299 (1993).
- Hoth, M. & Penner, R. Calcium release-activated calcium current in rat mast cells. *J. Physiol. (Lond.)* **465**, 359–386 (1993).
- Bakowski, D. & Parekh, A. B. Monovalent cation permeability and Ca^{2+} block of the store-operated Ca^{2+} current I_{CRAC} in rat basophilic leukemia cells. *Pflügers Arch.* **443**, 892–902 (2002).
- Lepple-Wienhues, A. & Cahalan, M. D. Conductance and permeation of monovalent cations through depletion-activated Ca^{2+} channels (I_{CRAC}) in Jurkat T cells. *Biophys. J.* **71**, 787–794 (1996).
- Prakriya, M. & Lewis, R. S. Separation and characterization of currents through store-operated CRAC channels and Mg^{2+} -inhibited cation (MIC) channels. *J. Gen. Physiol.* **119**, 487–507 (2002).
- Su, Z., Shoemaker, R. L., Marchase, R. B. & Blalock, J. E. Ca^{2+} modulation of Ca^{2+} release-activated Ca^{2+} channels is responsible for the inactivation of its monovalent cation current. *Biophys. J.* **86**, 805–814 (2004).
- Prakriya, M. & Lewis, R. S. Regulation of CRAC channel activity by recruitment of silent channels to a high open-probability gating mode. *J. Gen. Physiol.* (in the press).
- Voets, T. et al. CaT1 and the calcium release-activated calcium channel manifest distinct pore properties. *J. Biol. Chem.* **276**, 47767–47770 (2001).
- Yang, J., Ellinor, P. T., Sather, W. A., Zhang, J. F. & Tsien, R. W. Molecular determinants of Ca^{2+} selectivity and ion permeation in L-type Ca^{2+} channels. *Nature* **366**, 158–161 (1993).
- Ellinor, P. T., Yang, J., Sather, W. A., Zhang, J. F. & Tsien, R. W. Ca^{2+} channel selectivity at a single locus for high-affinity Ca^{2+} interactions. *Neuron* **15**, 1121–1132 (1995).
- Voets, T., Janssens, A., Prenen, J., Droogmans, G. & Nilius, B. Mg^{2+} -dependent gating and strong inward rectification of the cation channel TRPV6. *J. Gen. Physiol.* **121**, 245–260 (2003).
- Voets, T., Janssens, A., Droogmans, G. & Nilius, B. Outer pore architecture of a Ca^{2+} -selective TRP channel. *J. Biol. Chem.* **279**, 15223–15230 (2004).
- Eisenberg, B. Proteins, channels and crowded ions. *Biophys. Chem.* **100**, 507–517 (2003).
- Corry, B. & Chung, S. H. Mechanisms of valence selectivity in biological ion channels. *Cell. Mol. Life Sci.* **63**, 301–315 (2006).
- Feske, S., Giltner, J., Dolmetsch, R., Staudt, L. & Rao, A. Gene regulation by calcium influx in T lymphocytes. *Nature Immunol.* **2**, 316–324 (2001).
- Yeromin, A. V. et al. Molecular identification of the CRAC channel by altered ion selectivity in a mutant of Orai. *Nature* advance online publication, doi:10.1038/nature05108 (20 August 2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the National Institutes of Health to A.R. and S.F.; from the Charles H. Hood Foundation to S.F.; and a Scientist Development Award from the American Heart Association to M.P.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.R. (arao@cbr.med.harvard.edu).

LETTERS

An intron with a constitutive transport element is retained in a *Tap* messenger RNA

Ying Li¹, Yeou-cherng Bor¹, Yukiko Misawa¹, Yuming Xue¹, David Rekosh¹ & Marie-Louise Hammar skjöld¹

Alternative splicing is a key factor contributing to genetic diversity and evolution¹. Intron retention, one form of alternative splicing, is common in plants² but rare in higher eukaryotes^{3–8}, because messenger RNAs with retained introns are subject to cellular restriction at the level of cytoplasmic export and expression^{9,10}. Often, retention of internal introns restricts the export of these mRNAs and makes them the targets for degradation by the cellular nonsense-mediated decay machinery if they contain premature stop codons^{11,12}. In fact, many of the database entries for complementary DNAs with retained introns represent them as artefacts that would not affect the proteome¹¹. Retroviruses are important model systems in studies of regulation of RNAs with retained introns, because their genomic and mRNAs contain one or more unspliced introns¹⁰. For example, Mason–Pfizer monkey virus overcomes cellular restrictions by using a *cis*-acting RNA element known as the constitutive transport element (CTE)¹³. The CTE interacts directly with the Tap protein (also known as nuclear RNA export factor 1, encoded by *NXF1*), which is thought to be a principal export receptor for cellular mRNA¹⁴, leading to the hypothesis that cellular mRNAs with retained introns use cellular CTE equivalents to overcome restrictions to their expression¹⁰. Here we show that the *Tap* gene contains a functional CTE in its alternatively spliced intron 10. *Tap* mRNA containing this intron is exported to the cytoplasm and is present in polyribosomes. A small Tap protein is encoded by this mRNA and can be detected in human and monkey cells. Our results indicate that Tap regulates expression of its own intron-containing RNA through a CTE-mediated mechanism. Thus, CTEs are likely to be important elements that facilitate efficient expression of mammalian mRNAs with retained introns.

We noticed that several *Tap* (*NXF1*) cDNA GenBank entries from human, monkey and mouse retain intron 10 (accession numbers 62089415, 55731387, AK088634 and AK08652). In addition, two studies have described a *Tap* mRNA of 4.3 kb, a size consistent with complete retention of intron 10 (1,801 nucleotides), in addition to the 2.5-kb mRNA that represents the fully spliced mRNA^{15,16} (Fig. 1a). Because RNA with a retained intron might be restricted in its export and expression, we scanned the sequence of intron 10 for the presence of a 'cellular CTE' that could possibly overcome such a restriction. We found a region of ~100 nucleotides in the middle of intron 10 that is similar to the CTE sequence of Mason–Pfizer monkey virus (MPMV) and is highly conserved in human and mouse (Supplementary Fig. 1). By the MFOLD program¹⁷, this region was predicted to have a structure very similar to part of the previously verified structure of the MPMV-CTE¹⁸ (Fig. 1b). The upper stem-loop of the folded structure shows an almost identical secondary structure to the lower stem-loop of MPMV-CTE (Fig. 1c) with 38 conserved nucleotides. This sequence includes the loop region known to bind Tap. Whereas the MPMV-CTE contains two Tap-binding loops, however, the folded structure from *Tap* intron 10 has only one.

To determine whether the sequence from *Tap* intron 10 could function as a CTE to allow export and expression of an mRNA with a retained intron, we inserted the putative *Tap*-CTE into an HIV *GagPol* reporter construct that was similar to the one used to identify the MPMV-CTE¹³. This plasmid expresses HIV Gag and Pol from an mRNA that retains an intron, but only in the presence of a functional CTE^{18–20}. Although a CTE containing one Tap-binding loop is sufficient to mediate RNA export^{14,21}, efficient protein expression from viral reporters requires two Tap-binding sites¹⁸. We therefore transfected reporter constructs containing either a single or duplicated copy of the putative *Tap*-CTE into 293T cells, with or without plasmids expressing Tap and NXT1 because cotransfection with these plasmids is known to increase MPMV-CTE function^{20,22}. For controls, we transfected cells with plasmids containing either an irrelevant sequence (*U2*), the MPMV-CTE or an HIV *GagPol*-RRE reporter with or without a plasmid expressing HIV Rev, because the *GagPol*-RRE mRNA remains nuclear in the absence of Rev^{23,24}. In

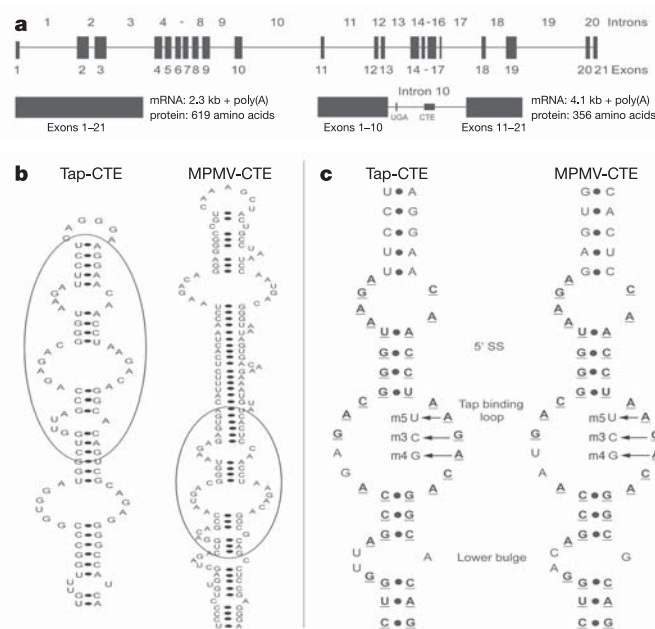


Figure 1 | Genomic organization of the human *Tap* gene and comparison of the *Tap*-CTE and MPMV-CTE. **a**, Genomic organization of human *Tap*. Vertical bars represent exons. The UGA stop codon and the putative CTE in intron 10 are indicated. **b**, Predicted secondary structure of the *Tap*-CTE and the MPMV-CTE structure¹⁸. **c**, Close up of the CTE structures (circled regions in **b**). Underlined bases indicate identity between the *Tap*-CTE and the MPMV-CTE. The m3, m4 and m5 mutations diminish Tap binding and CTE function in the context of the MPMV-CTE²⁵.

¹Myles H. Thaler Center for AIDS & Human Retrovirus Research and Department of Microbiology, University of Virginia, Charlottesville, Virginia 22908, USA.

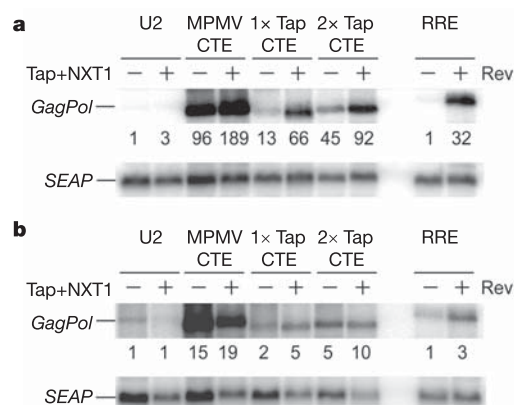


Figure 2 | Tap intron 10 contains a functional CTE. Northern blot analyses of cytoplasmic (a) or total (b) mRNA from 293T cells transfected with various GagPol constructs either with or without plasmids expressing Tap and NXT1. Shown are the fold differences in the levels of GagPol RNAs versus the GagPol-U2 RNA with or without Tap and NXT1 expression (left). As a control, a GagPol-RRE reporter was transfected with or without a plasmid expressing Rev (right). Values were normalized to SEAP mRNA expression from cotransfected plasmids.

addition, a plasmid expressing secreted alkaline phosphatase (SEAP) was used as a transfection and recovery control^{19,20}.

Northern blot analysis showed that transfection of one (1× Tap-CTE) or two (2× Tap-CTE) copies of the Tap-CTE increased the levels of cytoplasmic RNA 13- or 45-fold, respectively, as compared with the control U2 RNA (Fig. 2a). Transfection of plasmids expressing Tap and NXT1 caused an additional fivefold (1× Tap-CTE) or twofold (2× Tap-CTE) increase in cytoplasmic RNA. The differences in total RNA were much less than those observed for cytoplasmic RNA (Fig. 2b), showing that the Tap-CTE functions as an export element in conjunction with Tap and NXT1. The higher levels of GagPol total RNA observed with the CTE and RRE construct plus Rev probably reflect an increased rate of RNA degradation in the nucleus when no CTE or Rev is present. The Tap-CTE also mediated an increase in p24 expression, but only when duplicated or when Tap and NXT1 were also cotransfected (Supplementary Fig. 2). This might be due to additional translational functions of Tap that are required for retroviral expression²⁰. Thus, MPMV may have acquired its CTE from the Tap gene and subsequently duplicated it for more efficient viral gene expression.

We next examined endogenous Tap mRNA expression in 293T cells, with or without additional Tap and NXT1. Northern blot analysis, using a probe directed at exons 1–10 or intron 10, showed a species of roughly 4.3-kb in both total and cytoplasmic RNA (Fig. 3a). The exon probe also detected the completely spliced 2.5-kb mRNA and an overexpressed mRNA from the transfected Tap plasmid, which were not detected by the intron probe. Polymerase chain reaction with reverse transcription (RT-PCR) and DNA sequencing, using primers directed at intron 10 and exon 1, demonstrated that the 4.3-kb RNA retained intron 10 but completely lacked introns 1–9 (data not shown). Overexpression of Tap and NXT1 did not seem to change the ratios of the two endogenous Tap mRNAs or the levels of these RNAs in the cytoplasm, suggesting that 293T cells contain enough functional Tap to mediate export of the endogenous Tap mRNA containing intron 10.

To examine whether the intron-10-containing Tap mRNA that was present in the cytoplasm was associated with polyribosomes, we carried out sucrose gradient analysis on a cytoplasmic lysate from untransfected 293T cells. RNA isolated from individual fractions was analysed by northern blotting (Fig. 3b) or RT-PCR using an intron-10-specific primer pair (Fig. 3c). We found that Tap mRNA with a retained intron 10 was detected in the fractions containing polyribosomes, as was the fully spliced mRNA. This analysis also detected

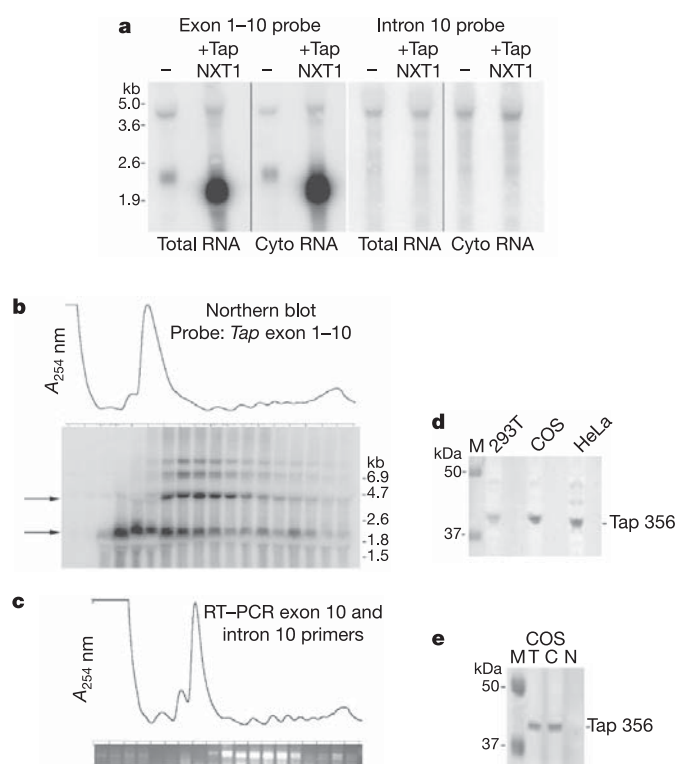


Figure 3 | Expression of Tap mRNA retaining intron 10 and a small Tap protein. a, Northern blot analysis of 293T cells using either exon-specific (left) or intron-specific (right) probes. Where indicated, cells were transfected with plasmids expressing Tap and NXT1. The darker bands on the left represent the Tap plasmid mRNA. b, c, Sucrose gradient analysis of cytoplasmic extracts. The gradient was scanned at an absorbance of 254 nm. RNA extracted from each fraction was analysed for Tap mRNA by northern blotting (b) or RT-PCR (c). Arrows in b indicate the 2.5-kb and 4.3-kb mRNAs. d, e, Western blot analysis of lysates from 293T, CMT3-COS and HeLa cells using an affinity-purified antibody raised against the peptide Ac-KRSVACVSPPGDLHPLGG-OH. M indicates molecular weight markers; T, C and N indicate total lysate, and cytoplasmic and nuclear extracts, respectively.

larger Tap mRNAs that might contain additional retained introns, but there is no evidence to indicate that there are additional CTEs. To confirm that the mRNA was present in polyribosomes, we carried out a similar experiment in the presence of 15 mM EDTA, a concentration that disrupts polyribosomes but not messenger ribonucleoprotein particle (mRNP) complexes (Supplementary Fig. 3). EDTA shifted both species of mRNA to 'lighter' fractions with, as expected, the mRNP complexes containing the larger mRNA sedimenting further down the gradient.

Because the polyribosome analysis suggested that Tap mRNA retaining intron 10 is actively translated in 293T cells, we analysed these cells and others for expression of a 'small' Tap protein. Tap mRNA retaining intron 10 would encode a 356-residue protein with 18 amino acids at its carboxy terminus encoded by the intron (Fig. 1a). Indeed, a protein of the expected size was detected in 293T, CMT3-COS and HeLa cells by using an antibody raised against this 18-residue peptide sequence (Fig. 3d). A small Tap protein was also detected in CMT3-COS cells, where fractionation of cells indicated that the protein accumulated mainly in the cytoplasm (Fig. 3e). A database search confirmed that the 18-residue peptide is unique to the Tap gene.

To study the functional role of the Tap-CTE in the expression of Tap mRNA, we transfected cells with a Tap minigene containing exons 1–10, intron 10 and exon 11, in parallel with a control cDNA construct that expressed a protein terminating at the in-frame UGA

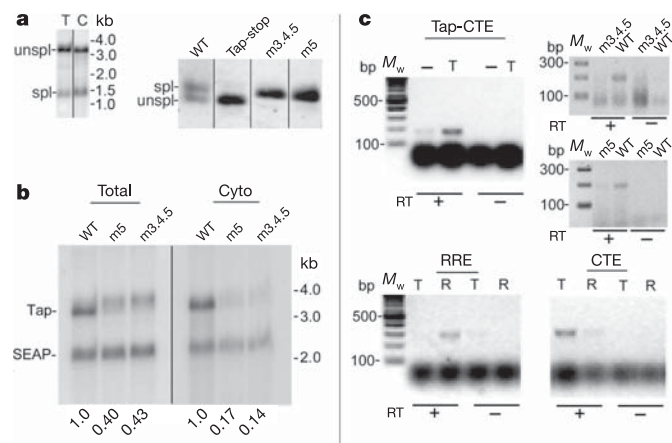


Figure 4 | RNA and protein expression from wild-type and CTE-mutated *Tap* minigene constructs. **a**, Northern (left) and western (right) blot analyses of 293T cells transfected with *Tap* minigene constructs. Total (T) and cytoplasmic (C) RNA expressed from the wild-type minigene construct was analysed using the exon 1–10 probe. Protein expression was analysed from cells transfected with indicated plasmids using a rabbit polyclonal antibody to Tap²⁰. **b**, Northern blot analysis of 293T cells transfected with the indicated constructs and a SEAP plasmid, using intron-10- and SEAP-specific probes. Numbers under each lane represent the fold amount of *Tap* mRNA relative to the wild type, after normalization to SEAP mRNA expression. **c**, Immunoprecipitation and RT-PCR analysis of ultraviolet-crosslinked RNA. Cells were transfected with the *Tap* minigene constructs (top), or the *GagPol*-RRE or *GagPol*-CTE (MPMV) reporters (bottom), and plasmids expressing Flag-Tap and NXT1 (T), Flag-Rev (R) or the empty vector (–). RNA was amplified using *Tap*-CTE- or *Gag*-specific primers. *M_w* indicates molecular weight markers; ‘+’ or ‘–’ indicates with or without reverse transcriptase (RT) in the PCR.

codon in intron 10 of *Tap* (Tap-stop). Northern blot analysis, using an exon probe, showed that the *Tap* minigene expressed two different mRNAs that were efficiently exported to the cytoplasm (Fig. 4a, left). The smaller species migrated at the position expected for the fully spliced mRNA, whereas the larger was the expected size for an RNA containing intron 10. The presence of the intron in this species was confirmed with an intron-10 probe (Fig. 4b). Western blot analysis (Fig. 4a, right) showed that two Tap proteins were expressed from the minigene. The smaller protein migrated to the same position as the Tap-stop protein, suggesting that it represents a protein translated from the intron-containing mRNA (356 residues). The larger protein was consistent with a protein translated from a completely spliced mRNA (367 residues). Only the smaller protein and Tap-stop was detected with the peptide antibody used in Fig. 3d, e (data not shown).

We also transfected cells with minigene constructs containing either single or triple nucleotide mutations in the *Tap*-CTE loop (Fig. 1c), which abolish Tap binding and function in the context of the MPMV-CTE²⁵. Both mutated constructs expressed only the larger Tap protein (Fig. 4a, right), suggesting that only the fully spliced mRNA reached the cytoplasm and was translated. Northern blot analysis (Fig. 4b) confirmed that the levels of intron-retaining cytoplasmic RNA from the mutated constructs were 6–7-fold lower than those from the wild-type construct. These reduced levels could not be due to RNA degradation, because there was only a twofold difference between the mutants and wild type in total RNA. Thus, the mutations clearly diminished export of the mRNA with the retained intron.

Tap mediates export of RNA containing the MPMV-CTE by binding directly to this element¹⁴. To determine whether Tap binds to the *Tap*-CTE in the context of *Tap* mRNA, cells were transfected with plasmids expressing Flag-tagged Tap and NXT1 and the wild-type or mutated *Tap* minigene constructs, in parallel with *GagPol*-RRE or

GagPol-CTE (MPMV) and Flag-tagged Rev or Tap. After ultraviolet crosslinking²⁶, cellular lysates were subjected to immunoprecipitation with anti-Flag beads. RT-PCR was done on RNA extracted from the precipitates, using either *Tap* intron-10-specific or *Gag*-specific primers. We detected a PCR product derived from the *Tap* mRNA (Fig. 4c), indicating that Tap could be crosslinked to the CTE. By contrast, very little PCR product indicative of crosslinking was observed with either of the mutants (Fig. 4c, top right). The specificity controls showed that Rev, but not Tap, crosslinked to the *GagPol*-RRE RNA (Fig. 4c, lower left) and that Tap, but not Rev, crosslinked to *GagPol*-CTE RNA (Fig. 4c, lower right). These results indicate that the binding of Tap to the CTE in intron 10 depends on the conserved loop sequence.

In conclusion, our results show that Tap regulates its own gene expression by binding to a *bona fide* CTE in intron 10, facilitating export and expression of mRNA containing this intron. Tap mRNA that retains intron 10 would be expected to be a target for nonsense-mediated decay (NMD), because downstream introns are removed^{11,12}. The presence of both the mRNA and protein in the cytoplasm suggests the possibility that export by the CTE somehow avoids the NMD machinery. Although the function of the small Tap protein is unclear, it retains the RNA-binding domain but lacks the nucleoporin and NXT1-binding domains, and locates mainly to the cytoplasm. Thus, it may regulate the function of Tap at the translational level²⁰.

These results point to a previously unknown mechanism of cellular gene regulation that facilitates the expression of alternatively spliced mRNAs with retained introns. *Trans*-acting factors, such as Tap and NXT1, would interact with a CTE to overcome the restriction at the level of nucleo-cytoplasmic export and expression. Such a mechanism could be potentially used in differentiation and development to regulate genes concertedly at the posttranscriptional level²⁷.

METHODS

More detailed methods are supplied in Supplementary Information.

Plasmids. Most of the plasmids used in this study have been described^{19,20,22}.

p24 ELISA and SEAP assays. Expression p24 from transfected 293T cells was determined by enzyme-linked immunosorbent assay (ELISA) using a p24 monoclonal antibody²⁸. SEAP activity in the supernatants was measured with the Phospha-Light Chemiluminescent Reporter kit (Tropix).

RNA crosslinking and immunoprecipitation. Cells were transfected with the plasmids described in Supplementary Information. They were then collected, crosslinked and lysed as described²⁶.

Polyribosome analysis, RNA fractionation, and northern and western blot analysis. These methods have been described^{19,20,22,29}. Our detailed step-by-step protocol for the northern blot analysis of RNA from mammalian polyribosomes is provided in ref. 30.

Received 6 May; accepted 24 July 2006.

- Kreahling, J. & Graveley, B. R. The origins and implications of Aluternative splicing. *Trends Genet.* **20**, 1–4 (2004).
- Ner-Gaon, H. *et al.* Intron retention is a major phenomenon in alternative splicing in *Arabidopsis*. *Plant J.* **39**, 877–885 (2004).
- Galante, P. A., Sakabe, N. J., Kirschbaum-Slager, N. & de Souza, S. J. Detection and evaluation of intron retention events in the human transcriptome. *RNA* **10**, 757–765 (2004).
- Ohler, U., Shomron, N. & Burge, C. B. Recognition of unknown conserved alternatively spliced exons. *PLoS Comput. Biol.* **1**, 113–122 (2005).
- Forrest, S. T., Barrington, K. G., Perleas, D., Hammarskjöld, M. L. & McNamara, C. A. Intron retention generates a novel Id3 isoform that inhibits vascular lesion formation. *J. Biol. Chem.* **279**, 32897–32903 (2004).
- Azios, N. G., Romero, F. J., Denton, M. C., Doherty, J. K. & Clinton, G. M. Expression of herstatin, an autoinhibitor of HER-2/neu, inhibits transactivation of HER-3 by HER-2 and blocks EGF activation of the EGF receptor. *Oncogene* **20**, 5199–5209 (2001).
- Bouillet, P. *et al.* Gene structure, alternative splicing, and chromosomal localization of pro-apoptotic Bcl-2 relative Bim. *Mamm. Genome* **12**, 163–168 (2001).
- Hiller, M., Huse, K., Platzer, M. & Backofen, R. Non-EST based prediction of exon skipping and intron retention events using Pfam information. *Nucleic Acids Res.* **33**, 5611–5621 (2005).

9. Chang, D. D. & Sharp, P. A. Messenger RNA transport and HIV rev regulation. *Science* **249**, 614–615 (1990).
10. Hammariskjöld, M.-L. Constitutive transport element-mediated nuclear export. *Curr. Top. Microbiol. Immunol.* **259**, 77–93 (2001).
11. Hillman, R. T., Green, R. E. & Brenner, S. E. An unappreciated role for RNA surveillance. *Genome Biol.* **5**, R8 (2004).
12. Lejeune, F. & Maquat, L. E. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr. Opin. Cell Biol.* **17**, 309–315 (2005).
13. Bray, M. *et al.* A small element from the Mason–Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. *Proc. Natl Acad. Sci. USA* **91**, 1256–1260 (1994).
14. Gruter, P. *et al.* TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Mol. Cell* **1**, 649–659 (1998).
15. Sasaki, M. *et al.* Molecular cloning and functional characterization of mouse Nxf family gene products. *Genomics* **85**, 641–653 (2005).
16. Herold, A., Klymenko, T. & Izaurralde, E. NXF1/p15 heterodimers are essential for mRNA nuclear export in *Drosophila*. *RNA* **7**, 1768–1780 (2001).
17. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
18. Ernst, R. K., Bray, M., Rekosh, D. & Hammariskjöld, M. L. Secondary structure and mutational analysis of the Mason–Pfizer monkey virus RNA constitutive transport element. *RNA* **3**, 210–222 (1997).
19. Coyle, J. H. *et al.* Sam68 enhances the cytoplasmic utilization of intron-containing RNA and is functionally regulated by the nuclear kinase Sik/BRK. *Mol. Cell Biol.* **23**, 92–103 (2003).
20. Jin, L., Guzik, B. W., Bor, Y. C., Rekosh, D. & Hammariskjöld, M. L. Tap and NXT promote translation of unspliced mRNA. *Genes Dev.* **17**, 3075–3086 (2003).
21. Pasquinelli, A. E. *et al.* The constitutive transport element (CTE) of Mason–Pfizer monkey virus (MPMV) accesses a cellular mRNA export pathway. *EMBO J.* **16**, 7500–7510 (1997).
22. Guzik, B. W. *et al.* NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of intron-containing RNA in mammalian cells. *Mol. Cell Biol.* **21**, 2545–2554 (2001).
23. Hammariskjöld, M. L. *et al.* Regulation of human immunodeficiency virus *env* expression by the *rev* gene product. *J. Virol.* **63**, 1959–1966 (1989).
24. Pollard, V. W. & Malim, M. H. The HIV-1 Rev protein. *Annu. Rev. Microbiol.* **52**, 491–532 (1998).
25. Kang, Y., Bogerd, H. P., Yang, J. & Cullen, B. R. Analysis of the RNA binding specificity of the human tap protein, a constitutive transport element-specific nuclear RNA export factor. *Virology* **262**, 200–209 (1999).
26. Ule, J., Jensen, K., Mele, A. & Darnell, R. B. CLIP: a method for identifying protein–RNA interaction sites in living cells. *Methods* **37**, 376–386 (2005).
27. Keene, J. D. & Lager, P. J. Post-transcriptional operons and regulons co-ordinating gene expression. *Chromosome Res.* **13**, 327–337 (2005).
28. Wehrly, K. & Chesebro, B. p24 antigen capture assay for quantification of human immunodeficiency virus using readily available inexpensive reagents. *Methods* **12**, 288–293 (1997).
29. Bor, Y. *et al.* The Wilms' tumor 1 (WT1) gene (+KTS isoform) functions with a CTE to enhance translation from an unspliced RNA with a retained intron. *Genes Dev.* **20**, 1597–1608 (2006).
30. Bor, Y.-c., Swartz, J., Li, Y., Coyle, J., Rekosh, D. & Hammariskjöld, M.-L. Northern blot analysis of mRNA from mammalian polyribosomes. *Nature Protocols* doi:10.1038/nprot.2006.216 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Morgenegg and F.-F. Chen for technical support, and J. Swartz, A. Ward and S. Raman for comments on the manuscript. This work was supported by grants from the National Institutes of Health (to M.-L.H. and to D.R.). Salary support for M.-L.H. and D.R. was provided by the Charles H. Ross Jr and Myles H. Thaler Endowments at the University of Virginia.

Author Contributions Y.L. and Y.-c.B. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.-L.H. (mh7g@virginia.edu).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

When Hurricane Katrina hit Louisiana and Mississippi a year ago, thousands of people lost their houses, possessions and even, tragically, loved ones. They also lost work. Many have relocated to cities such as Houston to continue with their professions or to start anew. But for some scientists, the task of re-engaging in their careers has been more complex. The flood waters claimed a host of data, along with valuable research materials and equipment. And for clinicians, the plummeting population of New Orleans has left them scrambling to find people who will participate in their studies (see *Nature* **442**, 856–857; 2006).

The anniversary of Hurricane Katrina serves as a sad reminder of the need for foresight in order to deal with unexpected disasters. One reader affected by the flooding asked for advice on how to handle the burdens of a busy assistant professorship, rebuilding his research and lab, and the desire to help his colleagues, students and family members cope with the aftermath. Our 'Ask an Expert' columnist advises that researchers deal with the tragedy much like any victim: make sure that you are coping properly before you start helping others; share your frustrations with fellow disaster victims; and regain control by finding a new routine, setting project milestones and adjusting your career timeline.

Relatively few researchers need worry about hurricanes. But changes in your personal life — the birth of a child, for example, or even an sudden illness — can also have an unexpected effect on your career (see page 240). For scientists who have coped with disasters such as Hurricane Katrina, the events serve as a reminder to make contingency plans. It's better to back up data before you lose them and to build some redundancies into experiments before your work is lost in an unforeseen incident. But the personal sphere should also be examined. What would you do if you needed to take a long leave? How would you cope if you unexpectedly lost a research partner or a job? Although no one knows when their life may change suddenly, you can anticipate that it might: hope for the best, and plan for the worst.

Paul Smaglik, *Naturejobs* editor

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
Natureevents:
Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestringuez (4994)

UK/Ireland/Italy/RoW:
Loredana Milanese (4944)
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

European Satellite Office
Germany: Patrick Phelan
Tel: +49 89 54 90 57 11
Fax: +49 89 54 90 57 20
e-mail: p.phelan@nature.com

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

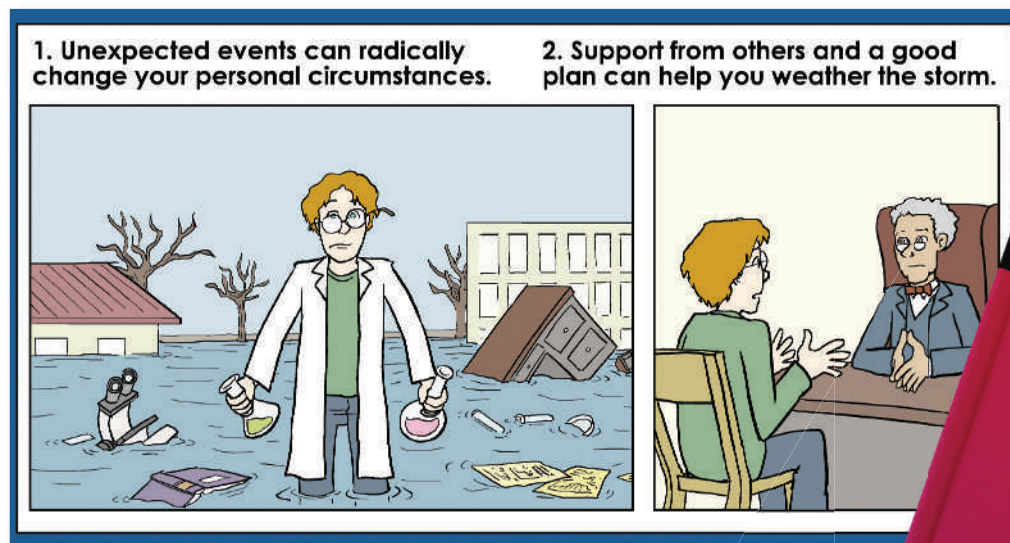
Naturejobs web development: Tom Hancock
Naturejobs online production:
Catherine Alexander

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com



When it's personal

Personal problems can overwhelm an already stressful young research career, but sound strategies can help you stay the course, counsels **Kendall Powell**.

When graduate student Xiujuan Zhang had her baby, Tate, on 7 August 2005, she knew that she needed to adjust the balance between time spent in and out of the lab. Three weeks later, both her research at Tulane University in New Orleans, Louisiana, and her life were turned upside-down by Hurricane Katrina.

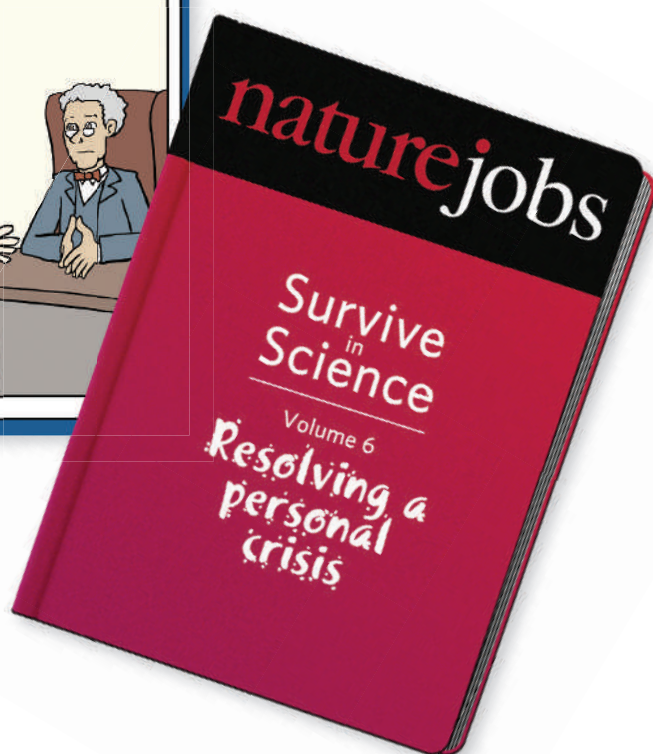
Zhang and her husband, both students from China, left the city with their newborn son, stayed with friends in Baton Rouge for a month, and eventually relocated to Houston, Texas, where she continued her research. In those trying times, Zhang kept her biomolecular-engineering PhD project alive by relying on her husband and her adviser, who were both supportive, and learning to compartmentalize her work and personal life. Her coping mechanisms kept her moving towards her research goals, even if at a much slower pace.

Disruptions to your personal life can take you away from the bench for significant chunks of time or cause ongoing distractions. Some issues simply require an adjusted work schedule. Others can be life-changing and, as a result, goal-changing. There are no one-size-fits-all strategies for dealing with rough patches that affect life both in and out of the lab. But researchers facing personal storms can draw on a mix of coping mechanisms, support networks and professional help.

Warning signs

The first step, say psychologists, is to recognize when a problem is overwhelming you. Warning signs of stress include physical symptoms such as stomach aches, headaches and insomnia, as well as emotional symptoms such as depressed moods, excessive worry and irritability, says Jeanne Manese, a psychologist in the counselling services of the University of California, San Diego.

"Often students at large research universities don't



want to let go of their optimal functioning" to deal with a problem, Manese notes. But isolating yourself and denying a problem are the worst things you can do, therapists say. Instead, lab workers should talk to someone caring yet objective.

"Early detection is always the best principle," says Darrell Smith, a psychologist in the biomedical trainee wellness programme at Vanderbilt University in Nashville, Tennessee. The sooner you seek help with a problem and get support systems going, the better the outcome is likely to be, he says.

The Vanderbilt programme rests on the principle that a counsellor who understands the research environment as Smith does will be better able to help scientists cope with crises. The programme encourages researchers to get help quickly, assessing the problem, and forming a plan that addresses both personal problems and research goals. If researchers need long-term professional counselling, Smith refers scientists to an outside specialist.

Riding out storms

Responses by various members of the Tulane group to the stressful situation highlight how greatly coping strategies vary among individuals. Zhang continued her research under tough circumstances because her husband was free to handle domestic affairs. Her adviser, W. T. Godbey, juggled both home and research duties, temporarily relocating to Rice University for research and making several trips between Houston and New Orleans. Colleague Shelly Corcoran devoted the full autumn semester to helping her family and her husband's family clean up and rebuild their hurricane-damaged homes; dealing with research too "would have been beyond overwhelming," she says.

Darrell Smith says it is important to seek help as early as possible.



After the group reunited in January, Godbey relaxed the reins a bit, giving lab workers time to cope in their own ways. He notes that some more heavy-handed colleagues had students who fled science altogether. Many scientists say that interspersing work with personal problem-solving helped them keep a better attitude, even if it meant slowing their productivity.

Now an instructor at a research university in the US northeast, Anne* went through a divorce and a diagnosis of a chronic neurological illness during graduate school. She kept her research and her personal struggles going in tandem “to keep some research momentum going, rather than focusing on the negativity surrounding me”, she says.

Likewise, Leila, whose mother was diagnosed with breast cancer last year, flew back and forth between taking care of her mother on the West Coast and doing her postdoctoral research in Boston. Because she had the support of her adviser, labmates and husband, Leila says she was able to “try to move towards my goal at the same time as taking care of my mum — I could work at my own, slower pace”.

Leila also leaned on her husband for emotional support, talking to him every day while she was away. When she returned to Boston, she found a counsellor in a cancer-support group. Talking to an impartial counsellor helped her dispel guilt, fear and anxiety.

Seeking help outside the lab first can give perspective about how to approach a supervisor. Marlena sought outside advice before telling her two advisers about a medical condition that would require three operations and prolonged recovery times. She got two great pieces of advice: be very clear about what you need from your adviser, and make a list of what research can proceed and what will have to wait.

Spilling the beans

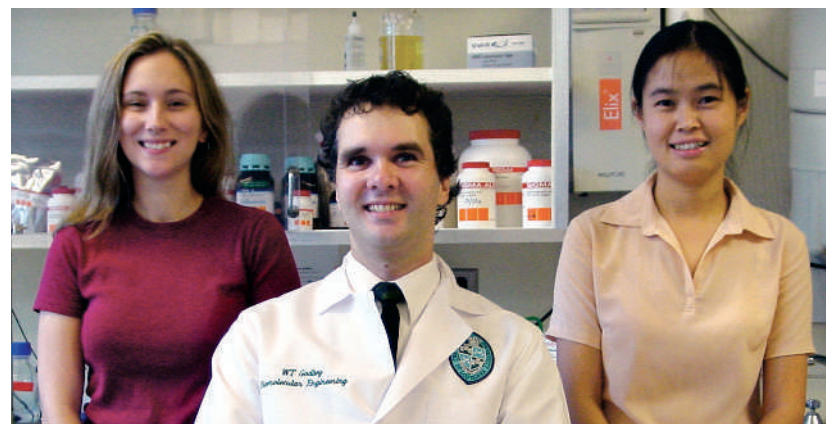
Making that list of research concerns “gave me something else to worry about, but it also gave my advisers confidence that I was going to come back to the project”, says Marlena. She gathered facts about her condition, determined how much time off she would need, then revised her research plans, with her advisers’ assistance.

“We tried to think of ways I could still be producing, but from the comfort of my couch,” she says. Marlena determined that a technician’s input and an undergraduate fieldwork assistant could keep her work on track. And while recovering at home, she could write the thesis introduction, review literature and work on a database project.



Jim Kendall says you shouldn't feel you have to tell everything to those around you.

W T. Godbey, centre (with Shelly Corcoran and Xiujuan Zhang), tells people to try not to give up their research during a stressful time.



Marlena’s advisers kept her on a fellowship during her leave, easing her financial burden. When asking an adviser for time off for family, medical or disability leave, it helps to know your rights as dictated by campus, department or funding-agency policies. If you have a rocky relationship with an adviser or anticipate that she will object to time off, gather information and allies first and go into the meeting prepared.

Minnie took two periods of leave during graduate school and a postdoctoral fellowship to help her grandmother, who raised her, to recover from surgery. In both cases, she explained why the leave mattered to her and that she would make up the time when she returned by working nights and weekends. Choose your battles wisely, Minnie says, with an adviser who can affect your entire career.

At the same time, it is important to remember that your adviser is not a therapist. “Don’t panic and feel you have to spill your guts about everything,” says Jim Kendall, manager of Vanderbilt’s employee assistance programme. “Sometimes less is better in the workplace environment,” he says, in case your lab head fears you won’t be able to function.

However, keeping problems entirely secret from your lab is usually not possible or advisable, says Smith. To ease the strain, Anne decided to tell her adviser about her divorce: “He needed to know why I was going to be distracted for the foreseeable future. To have a secret is just an added burden.”

Recovery time

Keeping lines of communication open during leave can maintain momentum. Godbey recommends also “keeping your hands wet” in the science. “To not work during a really trying time means you might make decisions differently,” he says. Godbey wishes he and Corcoran had kept in better touch during late 2005, to keep her connected to the temporary lab in Houston. After the hurricane, Corcoran decided to drop her doctorate in favour of a master’s.

After her second operation, Marlena had co-workers bring her into the lab for a few hours now and then, so that she could talk about science and attend seminars. “It helped me feel connected and made my advisers realize that I wasn’t just going to run away,” she says. Even when she could only handle television, she chose scientific documentaries on the Discovery Channel.

All these scientists say that keeping a meticulous lab notebook is critical for picking up where you left off. If you know you will be taking a break, note where you leave reagents, freeze down critical biological samples, and write up instructions for other lab members.

Keep your perspective about a temporary situation and don’t let it derail your long-term goals, crisis survivors say. “It would be a shame to throw away the schooling,” says Anne. “People don’t throw away other jobs for personal reasons.” Don’t beat yourself up or apologize for an unavoidable situation, either, but rely on other people for emotional and research support.

“You should not be afraid or ashamed to reach out,” says Leila, who was touched by her colleagues’ assistance during her mother’s illness. “This is not like an experiment you should do independently — you need help and you should let people help you.” ■

Kendall Powell is a freelance science writer in Broomfield, Colorado.

*Some names have been changed to protect privacy.

MOVERS

Richard Granger, director, Neukom Institute for Computational Science, Dartmouth College, Hanover, New Hampshire



1990–2006: Professor, University of California, Irvine

1980–89: Assistant rising to associate professor, University of California, Irvine

In 1971, when neuroscientist Hans-Lukas Teuber asked his class at Massachusetts Institute of Technology “How do we see a line as a line?”, Richard Granger’s mind blew a circuit. The first-year computer scientist became preoccupied with understanding how information was received and processed in the brain. He decided to refocus his career by integrating his long-standing love of computers with his new-found passion for unlocking the mysteries of the mind.

“A lot of the big questions are ones for which fields don’t yet exist,” says Granger. During his PhD in computer science at Yale, he designed computational tools to analyse psychological data related to fundamental aspects of memory — how to store, access and retrieve it — in order to gain insights into the neurobiological mechanisms at work. In 1980, as a young faculty member at the University of California, Irvine, he learned about brain circuitry by sitting through a graduate-level series of neuroscience courses. The primitive state of neuroscience as a field, the limited mathematical models and the sorry state of computers made him worry that delving into computational neuroscience was premature. But that didn’t dissuade him.

To circumvent these limitations, Granger built simplified, anatomically specific computer circuits to imitate real neurons from the bottom up. In 1990, while modelling the brain’s olfactory cortex (which was well understood anatomically), his circuit did something novel and surprising — it gave a series of meaningful responses to a single input. Since then, he has tried to tackle major questions such as how neurons interact and pass on messages to perform organism-level behaviours. In the past few years, Granger has received US patents in areas ranging from neurological diagnostics to financial analysis.

Once he heard of Dartmouth College’s new Neukom Institute for Computational Science, doing computation work in everything from scientific analyses to the creative arts, Granger knew the directorship was ideal for him. Carol Folt, Dartmouth’s dean of the faculty, agrees. She says Granger is one of the few people with the expertise and enthusiasm necessary not only to bring people from disparate areas together, but to help shape this new field.

Granger hopes to provide the spark to jumpstart the institute and ignite multiple collaborations. From basic work on perception to commercial applications of enhanced pattern recognition, he sees endless potential. The key, he says, is to look beyond what is known, to what is possible. ■

Virginia Gewin

RECRUITERS & ACADEMIA

Coming home

Returning home to Croatia after working as a postdoctoral fellow in the United States took preparation and patience. I had enjoyed my three years studying insulin resistance at two prominent US centres. Although I disliked the poor career prospects for postdocs and the lack of professional development, I loved my colleagues’ competitive spirit and the generous funding of the laboratory.

I always planned to return home, but the move required preparation because of Croatian hiring practices. Science is state-funded and scientists often stay where they trained, making institutions less open to postdocs from abroad. In an effort to retain professional contacts, I stayed in touch with my former professors via regular e-mails, visits during holidays and Christmas cards. When I started searching for jobs a year ago, I got two offers immediately. But it took ten months to secure a job. Positions first have to be approved by the government, then universities carry out a complicated protocol, followed by more government paperwork.

I now work as a research fellow and instructor. My monthly salary is about US\$900 after taxes, the laboratories are well equipped, and state funding is enough for respectable scientific production in terms of equipment,

papers and staff. Funding is limited, however. Science was hindered by war for the first 5 of Croatia’s 16 years of independence. And academic advancement requires a certain number of published articles, regardless of impact or importance.

There are problems with research logistics that I never encountered in the United States. Getting laboratory animals is mission impossible; on-site animal facilities are troublesome, and materials and equipment are expensive. Importing a \$10 rat from Italy costs \$200 for its transportation, quarantine and veterinary clearance.

Yet many people at the school have returned from abroad and started successful laboratories. Our new basic-science building is state-of-the-art. Fifteen years ago the school had no papers in respected international journals; now my department alone has two to four annually.

I came back to be an active part of the growth of Croatian science. Returning to a country in transition is both challenging and rewarding — though not recommended for gentle souls expecting a red carpet. ■

Livia Puljak, recently a postdoc at the University of Texas Southwestern Medical Center at Dallas, is now a research fellow at the University of Split Medical School in Croatia.

GRADUATE JOURNAL

Tables turned

It’s hard being a supervisor. When I was first told that an undergraduate student was coming to the lab, I thought, “Great. A slave!” Experiments would be finished quickly, coffee fetched diligently. I looked forward to instilling the joy of science into my new student, illuminating the path to a PhD. I would be the cool graduate student who shaped their scientific career.

But reality didn’t live up to my expectations. To begin with, I hadn’t realized how much teaching was involved, nor how often I would hear the words “What shall I do next?”. All the reading and other tasks I thought I could catch up on fell by the wayside as I had to supervise and then recheck different experiments. I started to dislike myself; I was selfish, intolerant and overbearing. I found it hard to let someone else be privy to my project; it was painful to watch others make mistakes. Often I had to bite my tongue or stop myself from taking over, reminding myself that I was in the same position once, eager to learn skills and needing to gain confidence.

Having experienced a short stint at supervising, my respect for my own PhD supervisor has grown immensely. How does she manage to be friendly but at the same time authoritative? How does she get us to realize our own mistakes and learn from them? She even looks as if she enjoys it. ■

Mhairi Dupré is a first-year PhD student in evolutionary developmental biology at the University of Oxford, UK.

The trial of Jeremy Owens

The memory lingers on.

Peter Roberts

The biological Jeremy Owens sat alone at the plaintiff's table. He had been rebuffed by every lawyer he had approached to assist him in this trial; some had laughed in his face. The ridicule only added to his determination to prove that he was the only real, true Jeremy Owens.

The judge entered the courtroom, rolled over to the bench, and made the necessary connections. Everyone was looking at the plaintiff. There was electricity in the air — or perhaps it was just the ozone generated by all the high-powered personages here to observe this high-profile trial.

The biological Jeremy Owens was, as usual, utterly confident. At first he had worried about receiving a fair trial — all the judges were, of course, enhanced. (Who wants justice rendered by mere human brains, so easily deceived, so weak in reason and judgment?) But he realized that the outcome of this trial depended solely on the justice of his claim. He knew he was right; how could he fail?

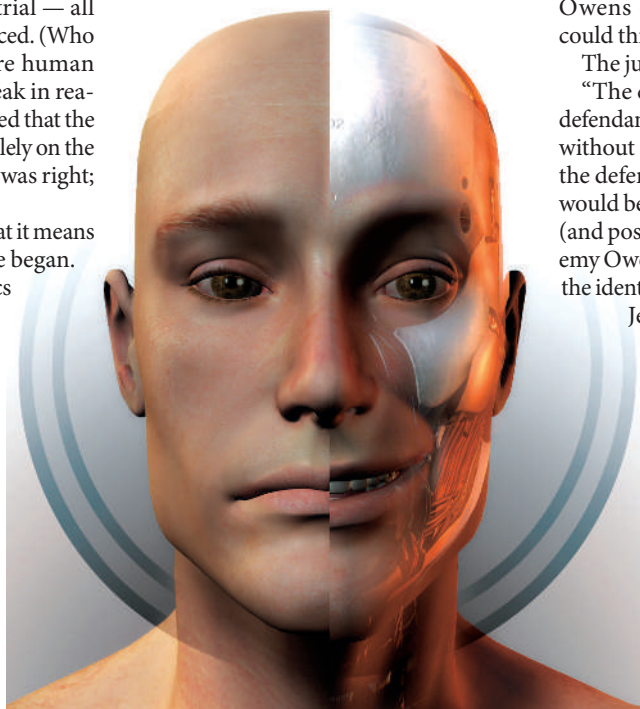
"I ask the court to consider what it means to be human, to be a person," he began. "It isn't simply a matter of genetics or physiology, defined by legalisms or high-flown philosophy. Rather, personhood consists of all the memories and experiences, thoughts, beliefs and feelings accumulated over a lifetime. I have experienced all the events of my life, whereas the defendant has not. Only I experienced my childhood, my school years, all the physical and emotional changes of growing up. Only I felt the confusion and thrill of my first love, and the trauma of its break-up. Only I suffered the skiing accident that sent me over a cliff and nearly ended my life. And only I have a right to my identity, my personhood. I alone am Jeremy Owens."

The response from the cyborg Jeremy Owens was immediate — and devastating:

"I, too, experienced that dreadful fall, and all the events and emotions of my tumultuous early life, for all the memories and thought patterns accumulated by that former appendage of mine at the plaintiff's table were downloaded and preserved, not to say treasured and learned from, in my extensive and very acute memory. If memory and experience are claimed as

the basis for identity, surely I must be the true possessor of that personhood, which you would claim for yourself? I remember everything you do, and much that you don't, and all of it in greater detail than you. Your memories are incomplete, partial, often faulty or false. Mine are complete, reliable, thorough and true.

"Who is more truly and fully Jeremy Owens? You, with your fuzzy, incomplete, vague recollections? Or me, possessed as I am of sharp, complete and full memories? You are a ghost, a pallid reflection of the body of experiences that forms the real Jeremy Owens. You are just a cast-off remnant, with no more claim on existence, on selfhood, than the thumb I had amputated earlier in my life."



"But, I've always been Jeremy Owens; you have not!"

"There has been a continuity and a trajectory from my birth — which produced that part of me now standing before this court as plaintiff — to the present day. The plaintiff is no longer a part of that trajectory, and for the court to recognize him as Jeremy Owens would disrupt the continuity of the life of the true Jeremy Owens, would diminish him, would make him less than he has already become, less than he is capable of being.

"When a snake sheds its skin, the cast-off skin has certainly been around longer

than the new skin, but it is not the snake. Shedding the old skin allows the snake to grow, to become more than it could have been had it retained — and been restricted by — its old skin. The plaintiff is like that discarded snakeskin: a cast off encumbrance, no longer a part of Jeremy Owens. Reinstating him as a part of Jeremy Owens could only diminish and constrain Jeremy Owens, preventing him from attaining his full potential. Surely, the court cannot permit such a miscarriage of justice. For the plaintiff even to claim to be a part of Jeremy Owens, let alone to claim, as he does, that he is Jeremy Owens, is a travesty. Therefore, we are confident that this court will rule in our favour."

Jeremy Owens — the biological Jeremy Owens — sat unmoving, stunned. He could think of nothing more to say.

The judge pronounced a verdict.

"The court finds the arguments of the defendant compelling, those of the plaintiff without merit. Therefore, we must rule in the defendant's favour. Furthermore, as it would be unthinkable to allow a discarded (and possibly defective) component of Jeremy Owens to wander the streets claiming the identity, rights and privileges of the real

Jeremy Owens, this court hereby enjoins the plaintiff from pretending to be Jeremy Owens; from using the name of Jeremy Owens in any way when referring to himself; from asserting any ongoing relationship to Jeremy Owens; or from claiming any of the rights or privileges of Jeremy Owens. The court further notes that it entertains grave doubts that the plaintiff is in any sense a person, or that he possesses any of the legal rights of a person. As the court has not yet

heard arguments on this issue, however, it would not be proper for it to hand down a ruling on that question."

Dumbfounded, devastated, the now nameless biological remnant of Jeremy Owens sat for a very long while. He was helped out of the building by a kindly officer of the court, who looked like nothing so much as a high-tech wheelchair. Jeremy was grateful for the assistance. ■

Peter Roberts has a BS in mathematics. His poetry and fiction have appeared in a variety of magazines over the past 30 years. See his website at www.geocities.com/peterroberts.geo.

JACEY